Abteilung für Allgemeinmedizin, Präventive und Rehabilitative Medizin

Geschäftsführende Direktorin: Prof. Dr. Erika Baum

des Fachbereichs Medizin der Philipps-Universität Marburg

Titel der Dissertation: Using multivariate techniques in the construction of a BMI risk index for Metropolitan Statistical Areas in the United States.

Inaugural-Dissertation zur Erlangung des Doktorgrades der gesamten Humanmedizin

dem Fachbereich Medizin der Philipps-Universität Marburg

vorgelegt von

Hans Kwame Kankam aus Goettingen

Marburg, 2009

Angenommen vom Fachbereich Medizin der Philipps-Universität Marburg am:

21.04.2010

Gedruckt mit Genehmigung des Fachbereichs.

Dekan: Prof. Dr. Matthias Rothmund

Referent: Prof. Dr. med. Annette Becker (MPH)

1Korreferent: PD Dr. Sitter

I. Overview

Overweight and obesity are the second leading causes of preventable death in the United States and are becoming major public health concerns on an international level (Yach 2006). Obesity has been defined as a pandemic by the World Health Organization (WHO) and is found to be prevalent among all strata of society, independent of ethnicity, demographics or socioeconomic status. This is alarming as the fundamental cause for this particular pandemic is far more complex in etiology and therefore more difficult to effectively battle than most epidemics that are caused by pathogens. The nexus of contributing factors is vast and is found to have physiological, psychological, economical and social roots, which all appear to interact and add to the rise in overweight and obesity among populations.

This study will review multilevel ecological and econometric modeling studies that focus on environmental and behavioral measures which exhibit significant association to overweight and obesity prevalence rates. The aim is to find variables that can be used as predictive markers in determining prevalence rates of overweight/obesity in nationally and internationally defined geographical units. The notion is to develop an index that is constructed from relevant covariates – gleaned from reviewed studies - that are found to significantly contribute to the cause of the rise in the prevalence of overweight and obesity (Body Mass). The covariates would be weighted in regard to their empirically evaluated contribution to the rise in prevalence of overweight and obesity in a population. This would provide a more powerful test of covariate balance than examining each covariate individually and could be used to predict the prevalence of overweight and obesity in communities such as metropolitan/urban areas and cities and rank them according to the populations' public health risk (in reference to the increased risk to be overweight or obese). For public health policy makers the importance of determining the strongest contributing factors to the prevalence of a disease process in geographical units is evident when it is necessary to assess health care requirements in the public health sector - for long-term planning in health care delivery, prevalence data are of key importance for budget planning and regulatory interventions. The decision to use multivariate techniques in the construction of an index is based on the fact that such a model has the benefit of offering a more in-depth analysis using specific subsets of survey data. Preliminary understanding of the reviewed data provides a feel for the meaning and distributional pattern of the data, indicating whether some variables have a greater variability than others (Everitt 2001). More research is being done in public health using multivariate techniques; this is based on the fact that survey and census data consist of data structures that are multivariate by their nature. Mandal et al (2006 & 2007) and Chang et al (2008) are among the researchers who have utilized multivariate methods in their statistical research and have demonstrated how these methods enhance the analyses of complex inter-correlations encountered in multilevel survey and census data. In this study we will utilize the multivariate technique in the construction of a risk index. The study seeks to demonstrate that we can analyze public health prevalence data and environmental measures that have multiple sources of variability and use the outcome to create an index that could determine a geographically defined population's risk to be overweight and/or obese.

II. **Obesity – Literature Review**

Overweight and obesity are measured by the body mass index (BMI), also termed Quetelet's index, and defined as weight in kilograms divided by height in meters squared (kg/m^2). Following national and international standards (World Health Organization, 1997; National Heart, Lung and Blood Institute, 1998), overweight is defined as BMI ≥ 25, obesity as BMI ≥ 30, class II obesity as BMI ≥ 35 and class III obesity as BMI ≥ 40.

Obesity has become the pandemic of the 21st century. In both developed and developing countries the percentage of overweight and obese individuals has been growing among all ethnic groups, socioeconomic levels and demographic segments. This is a cause of concern among global public health care systems, considering the mortality and morbidities associated with obesity; and the associated economic cost. It is estimated that in the United States alone the combination of diet, physical inactivity and tobacco are the leading causes of death, tobacco causing 435,000 deaths and diet & physical inactivity, 365,000 deaths in the year 2000 – of which Obesity is responsible for 112,000 deaths according to the Centers for Disease Control and Prevention (CDC).

In the U.S. 34 % of adults are considered overweight and an additional 31 % are obese, this equals to approximately seven out of ten adult Americans suffering from overweight and obesity. Among the younger population aged 2 through 19 years 16.3% were at or above the 95th percentile on the 2000 BMI-for-age growth charts and 31.9% were at or above the 85th percentile. This number is increasing annually as are direct and indirect health costs associated with the disease (Wolf 1998).

Overweight and obesity and their associated health problems have a significant economic impact on the U.S. health care system (U.S. Department of Health and Human Services 2002). According to a study of national costs attributed to both overweight (BMI 25–29.9) and obesity (BMI greater than 30), medical expenses accounted for 9.1 percent of total U.S. medical expenditures in 1998 and may have reached as high as $78.5 billion ($92.6

billion in 2002 dollars) with an additional estimated $40 billion in indirect costs (Finkelstein 2003). Direct medical costs as defined by the CDC – "include preventive, diagnostic, and treatment services related to obesity. Indirect costs relate to morbidity and mortality costs. Morbidity costs are defined as the value of income lost from decreased productivity, restricted activity, absenteeism, and bed days. Mortality costs are the value of future income lost by premature death."

There have been demands to implement far reaching public health intervention to counter the increasing prevalence of overweight and obesity (World Health Organization, 1997, World Health Organization, 2004, International Obesity Task Force, 2005). Up-to-date, interventions have been in the form of educational programs, to reduce overweight and obesity through healthy diet and physical exercise, which appear not to have been very effective considering the ongoing upward trend (Mandal B. C., 2006, Mandal B. a., 2007). When reviewing existing educational programs and information available to the public, we come across two pronounced information deficits. One concern is currently being addressed through legislation in New York City; to define and make publicly available the caloric information of food sold in limited-service and full-service restaurants as there is an established correlation between eating out and overweight and obesity (Chou 2004). The second and more challenging problem is to clarify what "appropriate weight" really means and what its definition is. To emphasize the issue, among the classified obese population 13 percent stated their weight to be in the norm or too low (Kuchler 2004). This misconception could either be caused by "health care illiteracy" in the population, or it could be the result of cultural differences - social norms. Norms tend to skew the perception of what is understood by individuals to be "normal weight" on the one hand and how it is defined by health organizations and providers on the other.

Overweight/obesity is a function of an individual's energy balance, which is measured as the difference between caloric consumption and caloric expenditure in a given period. Weight gain occurs whenever calories ingested exceed the calories expended through basic metabolism and activity; ergo, either a rise in calories ingested or a reduction in calories expended can lead to weight gain – overweight and obesity result from a positive imbalance in the energy equation, which leads to fat accumulation in the body (Loureiro 2005). It is not exactly clear to what extend this imbalance is caused by increase in calorie consumption or increase in sedentary lifestyle/decrease in physical activity. There is, however, a general consensus that overweight and obesity are caused by a multitude of contributing factors, occurring over an extended period of time. Fundamental energy balance factors such as gender, age, ethnicity, education, household income and socioeconomic status (SES) influence the process by which energy balances are translated into changes in Body Mass Index (BMI). The influence of genetics has been argued, but since our genes have not changed significantly during the past three decades this argument bears little weight. Other behavioral and environmental factors that have been considered as part of the cause of overweight and obesity include larger meal portion sizes, more snacking, and decreased smoking (Young 2002). Technological changes in food production and preparation as well as technological changes in our home and working environment have impacted our calorie consumption and expenditure as well as our leisure time activities (Bednarek 2006, Cutler 2003). The following section is a review of literature on the environmental and behavioral factors that have been researched and their contribution to the rise in prevalence of overweight and obesity in the US population.

1. Screen Time

Screen time is defined as time spent watching TV, playing electronic games and using the computer. A number of studies have found a positive relationship between sedentary lifestyles and overweight and obesity (Robinson, 1998, Jeffery, 1998, Kaur, 2003, Proctor, 2003). Screen time, which has been increasing among both adults and children, enforces sedentary behavior. Since the 1970s television has become the center piece of American leisure time - replacing available time for physical activity, consequently promoting a sedentary lifestyle. The amount of television watched was found to be positively related to the prevalence of obesity among children (Dietz 1986) and adults, (Tucker L. B., 1991, Tucker L. F., 1989). One specific study examined how television viewing and computer use are associated with BMI using NHANES 2003-2004 data (Doh 2007). The results suggest that adults who spent at least one hour per day on television viewing have BMIs that are 1.1 and 2.1 higher than adults who view television for less than one hour. For computer use the results indicated that adults who used computers for at least one hour had BMIs that are 0.5 to 0.9 higher than those who spent less than one hour on the computer. Although positive associations where found among both genders, there was a defined gender disparity. Effects on BMI were stronger in women for television viewing, and in men for computer use.

2. Technical Change

Some authors argue that technically induced change explains the growth of obesity during a period where there has been little increase in calorie consumption – indeed; there has been an increase in recreational exercise and dieting, which have been considered as substitutes brought about by technical changes in both market and household work (Cutler 2003). Even so, evidence proves the contrary to be true, despite off-the-job substitutes obesity has dramatically increased in the population.

The technical change hypothesis has been explained to have induced two specific changes.

- Lowered the cost of consuming calories by lowering the real cost of food (D. P. Lakdawalla 2002). Technological changes in agriculture reduced food prices consequently reducing market production costs of food (restaurants, fast-food and ready-to-cook-meals) which have substituted household production (Haines 1992).
- Raised the cost of expending calories by lowering the physical expenditure of calories per hour at work and at home – physical exertion when supplying labor has fallen (Cutler, 2003, Rashad, 2006)

The relative price of food has been declining constantly over several decades and is at a historical low – with the price of "unhealthy" food (calorie-dense foods) having declined more than healthy food (Gelbach 2007). It is estimated that there has been an annual decline of 0.2 percent in the real price of food from 1950 to 2000 (D. P. Lakdawalla 2002). In addition to the decline in food prices there has been a marked decline in time cost of food preparation which equates to a reduction in the per calorie cost of food (Cutler 2003). These effects can be noted in the prevalence of inexpensive calorie-dense convenience foods (snacks) and the dramatic growth of limited- service ("fast-food") restaurants. Chou et al, (2004), suggest that as much as 69 percent of the rise in BMI and 68 percent in the rise of obesity in the United States from 1978 to 2000 are related to the increase in the per capita number of restaurants. This outcome has been supported by other research, with limited-service restaurants being the main contributors (Kyureghian 2007). We should note that per capita number of restaurants could possibly be a proxy of technological change and increase in the value of time. It should be noted that Chou et al (2004) endeavored to investigate other factors that could have contributed to the rise of overweight and obesity between 1978 and 2000. They used individual level data from the Behavioral Risk Factor Surveillance System (BRFSS) and concluded that individual specific and demographic variables were not contributors to the rise in overweight and obesity during the given period. Technology induced sedentary job conditions and the dramatic decline in the relative price of food were researched by Lakdawalla et al (2002) using an econometric model, they estimated that these factors contributed 60 percent and 40 percent of recent weight growth respectively. Hill et al (2003) assert that their analysis of the National Health and Nutrition Examination Survey (NHANES) established an average weight gain in the population of 1.8 to 2.0 pounds per year in the 1990's. The findings postulate that the weight gained is associated with an increase of 15 to 30 calories consumed per day for median weight individuals and 50 to 100 calories per day for individuals at the 90th percentile. In a related study by Cutler et al (2003), a comparison of the 1977-1978 and 1994-1996 food intake surveys conducted by the U.S. Department of Agriculture showed similar findings. Men had increased their calorie intake by 268 calories and women by 143 calories over the given periods. This correlates with the rise in weight through the stated period. The researchers conclude that these calories come from calorie-dense, nutrient poor snacks throughout the day and not increased calorie intake through large restaurant meals and main meals at home.

3. Socioeconomic Status (SES), Education, Gender, Age, Race/Ethnicity

There is validity to the observation that overweight and obesity has a higher prevalence among the population with lower SES and lower education (Baum II, 2007, Amarasinghe, 2006). Women are found to have higher prevalence of obesity than men, African-American women having higher prevalence than Hispanic women who in turn have a higher prevalence than non-Hispanic white women. Women with lower SES are 0.5 times more likely to be obese irrelevant of their race and ethnicity. Men are about equally likely to be obese irrelevant of their SES, race and ethnicity, which ties into the fact that women experience negative effects on their economic outcomes with increase in body mass, whilst there is no such observed effect on men (Conley 2005), although this has been disputed in a recent study by Wada et al (2007). As people age they have a tendency to gain weight, this is observed in the general population - there is a quadratic relationship between age and overweight and obesity. With increasing age the SES-obesity gradient widens (Baum II 2007) consistent with studies that have analyzed health disparities among different socioeconomic groups. It should be noted that the rise in overweight and obesity has been uniform across all racial, socioeconomic and professional groups in the population over time and weight variation does not differ from other health conditions when comparing the rich to the poor (D. P. Lakdawalla 2002). This equal opportunity spread of overweight and obesity points to causes that affect all groups in the population. One could postulate that there has been a “trickle-down-effect” on all socioeconomic groups through rapid economic progress over time; this includes the negative and positive side-effects of economic prosperity in a society. One positive development with a backlash could be that in postindustrial and redistributive societies, including the United States, being unemployed is not a cause for weight loss, as food stamps and other welfare benefits are available to the indigent. Whether the Food Stamp Program (FSP) contributes to the rise in obesity has been debated in several studies, and has not been conclusive (Kaushal, 2007, Gibson, 2003).

4. Increase in the Value of Time

From the 1970s on women's participation in the labor force increased steadily. From 1970 to 2003 the number of employed adults increased by a factor of 1.8. Within the same time frame the number women in the workforce increased by the factor 5.1. This equates to an increase from approximately 30 percent female workers, of the total adult workforce (20 years of age and older) in 1970 to just below 50 percent in 2003 (Labor 1970-2003). The participation of more women in the workforce brought about innovative technological change in household production which was induced by the increased value of time. During this same time period overweight and obesity in children and adolescents started rising. Researchers have reviewed multiple factors that could have contributed to the increased trend in overweight and obesity in relationship to women's employment, especially maternal employment in association with child obesity (Anderson 2002). The evidence shows that there is a causal link to child obesity and maternal employment, although the magnitude of the effect is relatively small compared to the overall increase observed. This gives us reason to believe that other factors like, pre-cooked or ready-to-cook meals and food-away-from-home must have had a larger impact on rise in obesity. From 1970 onward technological change in the food industry enabled substitution of food-at-home meals with convenience foods, in parallel less time was spent in the household production process due to technical change in household appliances - these changes were driven by the increase in value of time, (Anderson, 2002, Lakdawalla D. P., 2005).

5. Built-Environment

As defined by Handy et al (2002) built environment consists of three interdependent factors;

- Urban design, the design of a city and its physical elements;
- Land use, location and density of residential, commercial, industrial, forest, and others;
- Transportation system, physical infrastructure of roads, sidewalks, bike paths, and others.

Urban-sprawl, also known as suburban-sprawl, is defined by neighborhoods that have low population density, poor accessibility to shops and workplaces, lack of sidewalks and generally poor access from one place to another (R. S. Ewing 2003). These limitations create an urban obesogenic environment, where travel for utilitarian purposes is done preferably by automobile, instead of walking or biking.

Research literature points to the fact that populations living in residential-sprawling neighborhoods are heavier than populations in less-sprawling and mixed-use neighborhoods. There is an ongoing debate between researchers on whether urban-sprawl contributes to the rise in obesity (R. S. Ewing 2003), or whether higher obesity rates in sprawling neighborhoods are due to self-selection of people who are predisposed to obesity (Eid 2007). For this study the “chicken or the egg” debate is of no importance. Evidence shows that there is a statistically significant difference in an overweight and obese population residing in sprawling neighborhoods compared to less-sprawling or mixed-use neighborhoods. All the studies reviewed controlled for demographic characteristics – gender, race/ethnicity, age, income and education – and all identified an association between overweight/obesity and urban sprawl, (Eid, 2007, Ewing R. B., 2006, Lopez, 2004, Papas, 2007). In his multilevel analysis Lopez (2004) used data from the Behavioral Risk Factor Surveillance System (BRFSS) in combination with an urban sprawl index. The index was developed by researchers at the Boston University of Public Health using the 2000 US Census data, sprawl values for 330 major metropolitan areas in the US were established on a scale of 0 to 100. The author concluded that the sprawl index scores were associated with increased risk of being overweight and obese - for each point increase in urban sprawl there was a 0.2 percent risk increase for being overweight and a 0.5 percent increase of being obese.

6. Smoking

The effect of cigarette smoking on the prevalence of overweight and obesity has been an interesting as well as controversial issue. The controversy exists mainly in the use of tobacco as a weight controlling agent among adolescent females (Cawley 2003). Both tobacco use and obesity are the top two public health issues associated with disease and premature mortality. Smoking is associated with lower weight and smoking cessation is associated with weight gain (Gruber 2005). This association has tempted us to make the general assumption that the successful fight to reduce the prevalence of smoking has had a direct effect on the rise in obesity. In a recent study by Flegal (2007) even large changes in the prevalence of smoking were shown to have little correlation to the rising prevalence of overweight and obesity. She concluded that decrease in the prevalence of cigarette smoking would probably not have a very significant effect, less than 1 percent, on the increase in the prevalence of overweight and obesity.

7. Proxy Connection - Technological Change, Relative Cost of Food, Increased Value of Time and the Built-Environment

After reviewing current literature on overweight and obesity, we proceed to determine which objective measureable variables could be proxies of other variables. We are aware from past research that individual specific and demographic variables are not important contributing factors to the steep rise in BMI (B. C. Mandal 2006) and that smoking cessation does not increase weight gain significantly in the long-run (Gruber, 2005, Flegal, 2007) which may be contrary to Chou et al (2004) whose data source was then used by Gruber et al. (2005) to come to opposing results. On the other hand, body weight is a function of individual characteristics and external factors such as sedentary life-style, technological change, the relative cost of food and increased value of time. In this section the focus is on the associations and correlations among variables that would lead us to representable objective variables, which could consequently be applied in calculating an overweight/obese risk index for a population in metropolitan statistical areas.

Lakdawalla and Philipson noted in their 2002 study, which looked at the rapid rise in BMI over three decades, that an increase in sedentary job conditions was a major cause of rise in weight in the initial period of 1971-1980, increase in sedentary job conditions stabilized between 1988-1994; NHANES II and III surveyed this period and a notable increase in BMI within 1988-1994 was observed. The authors postulated that the increase in BMI within 1988-1994 and NHANES 99 was caused by the defined decrease in food prices. The authors concluded that it was indicative of the influence of technological change in food production and technology based reductions in work-related energy expenditure and point to a direct association between technological change and lower food prices as well as sedentary market production (Lakdawalla D. P., 2002, Philipson, 1999, Lakdawalla D. P., 2005, Bednarek, 2006, Rashad, 2006). This period is noted for a shift in technological innovations on both the demand and supply side i.e. increase in food production, reduction in time spent on household and market production, and an increase in screen time; a defined period of economic progress.

Table 1 below shows a data summary from the National Longitudinal Survey of Youth (NLSY) developed by Lakdawalla et al (2002) from 1982-1998 showing the increase in BMI as the cohort ages with an increase in the prevalence of obesity. It also shows a decrease in job-related exercise (decrease in the level of strenuousness of jobs) in both genders over the same period of time. The NLSY is a survey conducted among a cohort of 12,686 people who were in an age range of 14 years and 22 years at the beginning of the survey in 1978. This cohort has been followed over time up to 1998, when the data was utilized by the authors. The results are suggestive to the association between technological change and reduced household and market production. It is an advantage that the NLSY maintains a consistent occupational coding scheme, contributing to the quality of the analysis.

Table 1: Summary of Statistics from NLSY 1982-1998 (Source: NLYS. 1982-1998)

	Working Men		**Working Women**	
	1982	1998	1982	1998
BMI	23.5	26.9	22.1	26.2
Obesity	0.05	0.22	0.05	0.27
Age	20.7	37	20.7	37
Black	0.14	0.14	0.14	0.14
Hispanic	0.06	0.07	0.06	0.06
Married	0.17	0.63	0.3	0.66
Highest Grade Attained	12	13.4	12.2	13.4
Distribution of Strength	%			
Strength=1	30.7	25.6	41.9	35.7
Strength=2	58.9	67.2	47.9	57.4
Strength=3	5.8	4.1	8.7	5.1
Strength=4	4.4	2.9	1.5	1.7
Strength=5	0.3	0.2	0	0.1
Distribution of Job-Related Exercise	%			
Strenuousness=0	43.5	37.3	43.6	39.8
Strenuousness=1	48.2	57	54.5	59.3
Strenuousness=2	1.9	0.3	0.4	0.4
Strenuousness=3	6.4	2.7	1.5	0.6

Strength is rated on a scale of one to five
Strenuousness is measured on a scale of zero to three

The authors utilized further datasets from the Dictionary of Occupational Titles, Fourth Edition, to ensure reduced errors in establishing decreases in job-related exercise – strenuousness. Reporting errors found in NLSY in relation to weights and heights were corrected using the National Health and Nutrition Examination Survey (NHANES) – as we find a reporting bias among both men and women who self-report weight and height.

Cutler et al (2003) propose a theory based on the division of labor in food preparation, which links back to the influence of technological change on the rise in BMI. The authors determined that in 1965, a female homemaker would spend in excess of two hours per day preparing meals – 57% of the total cost of food being time spent on preparation and clean-up. In 1995, the same meal preparation would take less than half that time. This change, they argue, was induced by the mass production and preparation of food, lowering the time and cost of food consumption. This change was enabled by technological innovations in food production and preparation; offering consumers easy and cost efficient access to ready-cooked-meals (88% of the cost of food eaten at home) and meals-away-from-home. The authors test their theory and gather empirical support. They show, as Lakdawalla et al (2002) do, that the consumption of mass produced food increased in the last two decades. The authors postulate that the decrease in the time cost of food preparation leads to an increase in food consumed; this is reflected in an increased intake frequency of convenience meals during the day, which encompass increased food varieties and high calorie/high flavor foods. Cutler et al test the various implications of their theory verifying the effect of technological change on increases in food production, preparation and consumption.

Table 2 below shows the time cost changes among the different demographic groups for meal preparations and cleanup comparing 1965 to 1995. The largest impact is observed among the group most affected by technological change in household production, married females that are not working.

Table 2: Time costs by demographic group (minutes). Source: Cutler et al calculations from Americans 'Use of Time Survey Archives, 1965 and 1995.

	1965		1995	
	Meal Prep.	Meal Prep. + Cleanup	Meal Prep.	Meal Prep. + Cleanup
Adults				
Single Male	13.6	18.1	15.5	17.3
Married Male, Non-Working Spouse	6.5	9.4	13.2	14.4
Married Male, Working Spouse	8.1	11.9	13.2	14.4
Single Female	38.1	60.1	28.9	33.1
Married Female, Working	58.3	84.8	35.7	41.4
Married Female, Not Working	94.2	137.7	57.7	68.8
Elderly				
Male	16.6	26.3	18.5	20.2
Female	65.9	10.4*	50.1	60.3

*A possible error by the authors

Table 3 shows the authors data on food consumption changes comparing 1977-1978 and 1994-1996, taking differences in meal consumption by gender into consideration. The data shows a clear increase in total calories consumed for both genders. The authors show in their results that the number of meals consumed per day has increased. The largest calorie increase was contributed by snacks, which fall under the category of convenience foods; these are inexpensive, highly processed food products requiring minimal preparation time that are extremely calorie dense and are easily available to consumers.

Table 3: Changes in Food Consumption, 1977-78 and 1994-96. Data are from the Continuing Survey of Food Intake 1977-78 and1994-96. *Average calories except for the row reporting average meals per day.

		Calories*			
	Meal	1977-78	1994-96	Change	Percent of total change
Male	TOTAL	2080	2347	268	100
	Breakfast	384	420	36	13
	Lunch	517	567	50	19
	Dinner	918	859	-59	-22
	Snacks	261	501	241	90
	Calories per meal	573	566	-7	
	Meals per day	3.92	4.53	0.61	
Female	TOTAL	1515	1658	143	100
	Breakfast	286	312	26	18
	Lunch	368	398	31	22
	Dinner	676	602	-74	-52
	Snacks	186	346	160	112
	Calories per meal	422	408	-14	
	Meals per day	3.86	4.44	0.58	

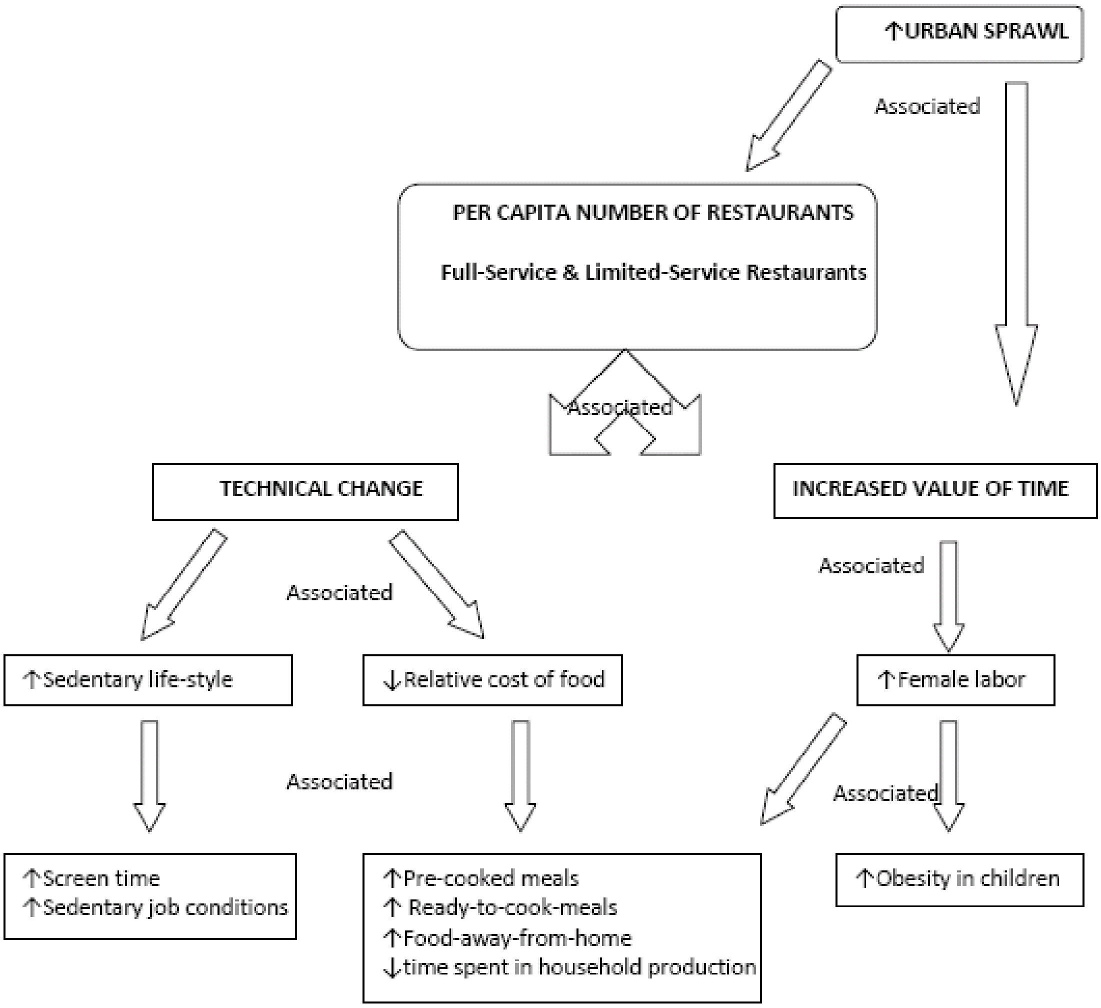
The argument is that there exists an association between per capita number of restaurants, technological change and relative value of time; where per capita number of restaurants could be the proxy for the variables. It is known that BMI has a high positive elasticity in respect to per capita number of restaurants (Chou 2004). Joanne Guthrie et al. (2002) estimated that 17 percent of total calories consumed by men aged 18 to 39, and 13 percent of the total calories consumed by women in that age group came from fast-food establishments. Growth in restaurants, particularly fast-food restaurants, has been dramatic in the past three decades so has the shift away from the consumption of meals at home (Kyureghian 2007). According to the Census of Retail Trade, the per capita number of fast-food restaurants doubled between 1972 and 1997, while the per capita number of full-service restaurants rose by 35% (Bureau of the Census, 1976, 2000). Recent research by Dunn (2007) as well as Mandal (2006) look at the relation between availability of fast-food and increase in BMI and come up with similar results as Chou et al., indicating the significant impact restaurants have on increase in BMI, confirming the association. Fast-food and convenience foods have common characteristics, they are inexpensive and have a high caloric density to make them palatable, another important attribute they share is easy and quick availability; this reflects on the relative value of time. The increases in hours worked and labor force participation rates, and declines or modest increases in real income experienced by certain groups appear to have stimulated the demand for these easily accessible foods. This gets compounded by the reduction in the time available for active leisure, resulting in reduced calories expended, (Lakdawalla D. P., 2005, Philipson, 1999).

It can be argued, from reviewed research that in areas or locations where demand for convenience foods is low due to the community characteristics, there will be less per capita number of restaurants and vice-versa; a neo-classical economic model of attaining supply and demand equilibrium. Restaurants do not open randomly; there is a rational decision by an entrepreneur to open outlets in the most profitable locations. This is further explained by Chou et al who argue that a meal in a restaurant requires travel and waiting time, which would be reflected in the price of a meal and would be an indirect measure of wage rates or hours of work. Consequently more restaurants would be found in areas where consumers have relative high time values, inherently reducing travel and waiting time. This would infer that there will be a larger number of fast-food restaurants than full service restaurants due to less waiting time. The authors included the per capita number of fast-food and full-service restaurants in their empirical analysis. There could be cause for endogeneity when restaurants are used in a regression as a variable, this potential problem was addressed by Chou et al who established that rise in per capita number of restaurants began before rise in BMI; a point of importance as the index we want to create has to be determined by objective verifiable variables and endogeneity would create a bias in the regression coefficient. From evaluation of past research it can be concluded that the per capita number of restaurants is associated with, and would be a viable proxy measure for technological change, the relative cost of food and increased value of time and could be used in calculating the BMI risk index. Part of the analysis is to evaluate whether there is any significant difference when testing for total number of restaurants as one variable or for fast-food and full service restaurants as two separate variables.

In their longitudinal analysis of the prevalence of obesity in the developed world, Bleich et al (2007) find that apart from the effect of technological change other sociodemographic factors including increased urbanization are found to have a positive association to increased rise in BMI. The association between urban sprawl and the risk of rise in BMI has been researched by Ewing et al (2003), Lopez (2004), Eid et al (2007), and Ewing et al (2006), all concluded that people living in sprawling neighborhoods have a higher BMI than people living in less sprawling neighborhoods. Although Eid et al contests any causal link between higher BMI and urban sprawl, they do concede to the positive association, which is of essence in this study as discussed earlier. Based on research by Lopez, where he concluded that for each point increase in urban sprawl there was a 0.2 percent risk increase for being overweight and a 0.5 percent increase of being obese urban sprawl is chosen as another associated variable that will be used as a contributor to our risk index.

This study analyzed variables that could be proxies for leisure time physical activity; measures such as the per capita number of gyms, access to recreational areas, parks and other green areas within a defined community were assessed. Reason for inclusion was based on the hypothesis that increases in physical activity, which could be indirectly measured by these variables, would be negatively correlated to rise in BMI. The Geographical Information System, (GIS) would be the tool of choice. It gives information on the availability of recreational facilities and explores the distribution of these facilities by population characteristics. Eight Standard Industrial Classification (SIC) codes for physical activity facilities and resources are used that fall within specified zip codes. When analyzing the Metropolitan Sprawl Index developed by Smart Growth America (SGA) it was discovered that these facilities and green areas had been taken into consideration and integrated into the sprawl index by one of the 4 factors, which is defined by SGA as "Strength of activity centers and downtowns". It was concluded that the sprawl index developed by SGA would be a proxy variable for built-environment; which includes green areas, recreational facilities and parks.

Figure 1: Association between Variables (arrows show increases and decreases)



III. **Methodology**

1. Model

The reviewed studies have made great steps towards identifying factors that contribute significantly towards the rise in body mass across the population. This ecological study could be considered a continuation of the studies reviewed. This study will adopt appropriate datasets and use multivariate methods, to create a BMI risk index for Metropolitan Statistical Areas (MSAs). The suitable model to fit the data will be a single-effect and interaction-effect multivariate/linear regression model with an intercept. A hierarchical linear model (HLM) could be used to account for potential violations of the independence assumptions, but is deemed unnecessary as we do not have to control for race, ethnicity, age, or other variables at the MSA level in our final model (Chou 2004). Chou et al found in their research that individual specific and demographic variables where inherently not contributing factors to the rise in overweight and obesity.

2. Data

The reviewed literature helps us decide on the individual-level and higher-level datasets most appropriate for this study. Since BMI is dependent upon reports or measurements of respondent height and weight, and since it is the main focus of this study in creating a BMI risk index at the MSA level in the United States, the study is limited in the dataset selection to those publicly available datasets that provide measurements of height and weight. The publicly available datasets containing measurements of height and weight at the MSA level are the datasets provided by the US Centers for Disease Control and Prevention's Behavioral Risk Factor Surveillance System (BRFSS), these data were selected for the analysis. Higher-level data is obtained from the Census of Retail Trade (2002) for the restaurant variables, population data for the MSA are obtained from US Bureau of the Census and from Smart Growth America (SGA), (R. P. Ewing 2002) the Metropolitan Sprawl Index (MSI) which represents the variable for built-environment in MSAs was adopted. It is important to take a detailed look at these datasets, their sources and why they are chosen for this study. The complete table of data used is found in Appendix 1.

a. Metropolitan Sprawl Index

The Metropolitan Sprawl Index (MSI) was developed by SGA in 2002. SGA defined sprawl as;

"*Low-density development with residential, shopping and office areas that are rigidly segregated; a lack of thriving activity centers; and limited choices in travel routes*."

The MSI is currently the most precise measure of urban-sprawl available and is considered the first index that takes into consideration the multidimensional characteristics of urban-sprawl. It is a linear combination of 22 land use and street network variables that are grouped into 4 factors that classify sprawl. In contrast to MSI, other urban-sprawl indices that have been developed used only one or two variables in defining sprawl; mainly population growth and land use, leaving out the highly complex effects on a wide variety of quality-of-life indicators. The 4 factors that were measured and analyzed in the MSI are;

- Residential density
- Neighborhood mix of homes, jobs and services
- Strength of activity centers and downtowns
- Accessibility of the street network

These 4 variables are further composed of 22 measurable components listed in **Table 4** below;

Table 4: Variables used to define sprawl (Source: Smart Growth America)

Factor	Variable	Source
Residential Density	Gross Population Density in persons per square mile	US Census
	Percentage of population living at densities less than 1,500 persons per square mile (low suburban density)	US Census
	Percentage of population living at densities greater than 12,500 persons per square mile (urban density)	US Census
	Estimated density at the center of the metro area	US Census
	Gross population density of urban lands	USDA Natural Resources Inventory
	Weighted average lot size for single family dwellings (in square feet)	American Housing Survey
	Weighted density of all population centers within a metro area	Claritas Corporation
Neighborhood Mix of Homes, Shops and Offices	Percentage of residents with businesses or institutions within 1/2 block of their homes	American Housing Survey
	Percentage of residents with satisfactory neighborhood shopping within 1 mile	American Housing Survey
	Percentage of residents with a public elementary school within 1 mile	American Housing Survey
	Balance of jobs to residents	Census Transportation Planning
	Balance of population serving jobs to residents. Population serving jobs include retail, personal services, entertainment, health, education, and professional services	Census Transportation Planning
	Mix of population-serving jobs	Census Transportation Planning
Strength of Metropolitan Centers	Variation of population density by census tract	US Census
	Rate of decline in density from center (density gradient)	US Census
	Percentage of population living within 3 miles of the central business district	Edward Glaeser, Brookings Institution
	Percent of the population living more than 10 miles from the CBD	Edward Glaeser, Brookings Institution
	Percentage of the population relating to centers within the same metropolitan statistical area	Claritas
	Ratio of population density to the highest density center in the metro area	Claritas
	Average block length in urbanized portion of the metro area	Census TIGER files
	Average block size in square miles	Census TIGER files
	Percentage of small blocks	Census TIGER files

The MSI was evaluated for 83 U.S. metropolitan areas that contained the complete dataset. These evaluated MSAs represent the largest metropolitan areas in the nation and represent nearly half of the U.S. population.

Other datasets were adjusted to the chronological development of the MSI by Ewing et al in 2002. We therefore used the BRFSS 2002, NHANES 2002, Census of Retail Trade 2002 and US Bureau of the Census 2002 to obtain our remaining variables.

b. BRFSS 2002

The BRFSS is an annual telephone survey of adults; there are 247,964 records for 2002. In 2002, there were 98 Metropolitan/Micropolitan Statistical Areas (MMSA) that had at least 500 respondents in the 2002 BRFSS and at least 19 sample members in all final weighting classes. The survey has more than 200 self-reported and calculated variables and is a good source of information about the health status and habits of the US population. Within each state, blocks of 100 telephone numbers constitute a primary sampling unit, and the goal is to interview 3 households within each primary sampling unit. The data are weighted on the basis of the probability of the household's telephone number being selected and the number of adults and telephones within the household. The sample is also adjusted for non-responses and for households without telephones. Finally, there is an adjustment to ensure the sample contains representative proportions of selected demographic characteristics (gender, age, and race/ethnicity). The BRFSS excludes institutionalized persons, and response rates vary by state. Data from the 2002 survey were obtained from the BRFSS Web site.

There are limitations to BFRSS data that have to be mentioned and eventually corrected;

- BRFSS generally oversamples less populous states
- Self-reported anthropometric variables are underreported and have to be adjusted using actual physical measurements from NHANES.

In this study it is necessary to correct for the self-reported anthropometric variables using NHANES 2002.

c. 2002 Economic Census of Retail Trade

The 2002 Economic Census of Retail Trade (Bureau of the Census 2002) presents data based on the North American Industry Classification System (NAICS), the restaurant data classification for full- and limited-service restaurants is defined as follows;

NAICS 722110 Full-Service Restaurants:

This industry comprises establishments primarily engaged in providing food services to patrons who order and are served while seated (i.e. waiter/waitress service) and pay after eating. These establishments may provide this type of food services to patrons in combination with selling alcoholic beverages, providing carry out services, or presenting live nontheatrical entertainment.

NAICS 722211 Limited-Service Restaurants:

This U.S. industry comprises establishments primarily engaged in providing food services (except snack and nonalcoholic beverage bars) where patrons generally order or select items and pay before eating. Food and drink may be consumed on premises, taken out, or delivered to the customer's location. Some establishments in this industry may provide these food services in combination with selling alcoholic beverages.

The data was obtained for all payroll and non-payroll restaurants in the MSAs classified by the SGA sprawl index.

d. Metropolitan/Micropolitan Statistical Areas (MMSA)

Data on the population in metropolitan areas for 2002 was obtained from the U.S. Census Bureau on both MSAs and primary metropolitan statistical areas, or PMSAs. As noted by SGS, “PMSAs are generally larger than political jurisdictions such as cities, but smaller than the entire metropolitan region; some regions may include several PMSAs, which are then combined to form a Combined Metropolitan Statistical Area (CMSA)". The population numbers we used were adapted to fit the data from the MSI. The sprawl index includes every metro over 500,000 population for which SGA could obtain a complete dataset.

3. Selected Variables

In the preceding section II under heading 7, (page 17) the various variables were analyzed to determine proxy-variables that could be used in the statistical study. Variables were selected that had the most impact on social- and work-environments over the period of 1971-2003, and were independent contributors to the rapid increase in community BMI according to prior research. The variables are indicators of individual and external characteristics; sedentary life-style, reduced household and market production, technological change, the relative cost of food, increased value of time and built-environment. The selected variables are per capita number of limited-service restaurants, per capita number of full-service restaurants, per capita number of total restaurants and urban-sprawl. Individual level data was obtained from the BRFSS and NHANES whilst the sprawl index of our choice representing built-environment is the MSI index developed by SGA. **Figure 1** (page 24) demonstrates, using a chart, how to associate and connect key variables to determine proxy-variables, which can be independently associated with weight status in the statistic. The resulting proxy-variables are per capita limited-service restaurants, per capita full-service restaurants and urban-sprawl. Individual level data will be obtained from the BRFSS and NHANES whilst the sprawl index representing built-environment is the index developed by SGA.

4. Develop Adjustment Weights

When selecting the BRFSS as the data source for this study, it was known that the data was limited to only self-report of height, weight, and BMI. This was a serious drawback to the investigation that had to be corrected as it has been widely reported that self reports of respondent height and weight are often fraught with measurement error and bias, with heavier people tending to underreport their weight. In order to counter this, the focus is on developing a valid statistical model that, when provided with self-reports of a respondent's height and weight, would provide more reliable estimates of the respondent's actual height and weight. To develop this BMI adjusting model the 2002 National Health and Nutrition Examination (NHANES) survey was used, which contained both respondent reports of height and weight as well as actual measurements of height and weight which were obtained shortly after collection of the self reports.

After downloading the NHANES data and reading it into the SAS System Version 9.2, 16 BMI adjustment models as suggested by Cawley (1999) were developed. Models were developed separately for weight and height for each of eight groups based upon the combination of race/ethnicity and gender. To develop the adjustment model for each group, the known quantity of interest [either height or weight (BMXWT/BMXHT)] was regressed upon the self reported value of the quantity (WHD020/WHD010) and its squared value (WHD020_sqrd /WHD010_sqrd). In each linear regression model developed, the intercept was forced to be zero, so the resulting models were represented by:

$$Y = \beta_1 X_1 + \beta_2 X_1^2$$

Where

β_1= the coefficient or adjustment weight associated with the self-reported quantity of interest;
X_1= the self-reported quantity of interest; and
β_2=the coefficient or adjustment weight associated with the squared value of the quantity of interest.

Another, less mathematically verbose model representation is:

$$Predcited\ Actual\ Value = \beta_1(Self\ Reported\ Quantity$$
$$+ \beta_2(Self\ Reported\ Quantity)^2$$

The summary results of each of the regression models are provided in **Tables 5** and **6** as well as the name of the results table containing more detailed model statistics:

Table 5: Adjustment Model for Weight (kg)

Adjustment Model for Weight (kg)					
	β_1 ***t*-statistic** ***p*-value**	β_2 ***t*-statistic** ***p*-value**	**Mode l N**	**Model MSE** **F-statistic** **p-value**	**Results Table**
Male, Non-Hispanic White	0.99855 173.05 <.0001	-0.00000850 -0.14 0.8864	1233	16.01759 301363 <.0001	Table 12: NHWM
Female, Non-Hispanic White	1.04676 135.38 <.0001	-0.00019439 -2.05 0.0405	1329	24.76330 150254 <.0001	Table 13: NHWF
Male, Non-Hispanic Black	0.93083 77.42 <.0001	0.00064095 5.19 <.0001	518	30.02721 65639.9 <.0001	Table 14: NHBM
Female, Non-Hispanic Black	1.04576 97.84 <.0001	-0.00022731 -1.98 0.0484	537	31.21623 58437.9 <.0001	Table 15: NHBF
Male, Hispanic	0.94894 87.50 <.0001	0.00058188 4.81 <.0001	628	26.12904 79810.3 <.0001	Table 16: HM
Female, Hispanic	1.04709 87.28 <.0001	-0.00026799 -1.75 0.0802	662	26.65025 64954.9 <.0001	Table 17: HF
Male, Other Race	0.95612 50.40 <.0001	0.00059623 2.75 0.0074	78	10.98299 21815.0 <.0001	Table 18: MRM
Female, Other Race	1.01385 28.83 <.0001	0.00042230 0.84 0.4016	99	25.71952 8421.00 <.0001	Table 19: MRF

Table 6: Adjustment Model for Height (cm)

Adjustment Model for Height (cm)					
	β_1 *t*-statistic *p*-value	β_2 *t*-statistic *p*-value	Mode l N	Model MSE F-statistic p-value	Results Table
Male, Non-Hispanic White	1.07720 85.66 <.0001	-0.00047969 -6.82 <.0001	1233	10.58422 1816679 <.0001	Table 20: NHWM
Female, Non-Hispanic White	1.07449 106.03 <.0001	-0.00048268 -7.81 <.0001	1329	6.38829 2754688 <.0001	Table 21: NHWF
Male, Non-Hispanic Black	1.18710 62.46 <.0001	0.00010608 -10.25 <.0001	518	10.90689 745448 <.0001	Table 22: NHBM
Female, Non-Hispanic Black	1.19083 62.21 <.0001	0.00119 10.27 <.0001	537	9.76295 730785 <.0001	Table 23: NHBF
Male, Hispanic	1.26969 65.75 <.0001	-0.00160 -14.23 <.0001	628	15.84437 575999 <.0001	Table 24: HM
Female, Hispanic	1.33293 57.39 <.0001	-0.00213 -14.64 <.0001	662	19.33825 424877 <.0001	Table 25: HF
Male, Other Race	1.11543 32.08 <.0001	-0.00068569 -3.41 0.0011	78	5.10194 225231 <.0001	Table 26: MRM
Female, Other Race	1.03838 27.17 <.0001	-0.00028159 1.18 0.2420	99	6.41349 193068 <.0001	Table 27: MRF

In every case, the adjustment models developed were statistically significant, being highly predictive of the actual values of height and weight.

IV. Results, Discussion and Conclusion

1. Develop Mean BMI's per MSA

Having built statistical models to adjust weights and heights for their underreporting, the next step was to apply the models to the appropriate demographic subsets in the 2002 BRFSS data. For each gender and racial group in the BRFSS data, the corresponding adjustment model was applied, multiplying the self reported quantities of interest and their squares by their respective weights to obtain the adjusted values. The adjusted values for heights (`PredictedHeight`) were then scaled from centimeters to meters by dividing by 100. The resulting values were then substituted into the formula for Body Mass Index:

$$BMI = \frac{weight\ in\ kilograms}{height\ in\ meters^{\ 2}}.$$

Once the adjusted BMI values were calculated (`BMI_adj`), summary estimates of the mean BMI per Metropolitan Statistical Area were obtained using the SAS surveymeans procedure which properly calculates the means of the BMIs for each MSA given the weighted values and complex survey sample design of the BRFSS study. The mean BMI values for each MSA are presented in the table below along with their lower and upper 95% confidence intervals and the number of observations used in the calculation:

Table 7: Adjusted BMI (N=111,704)

MSA Name (CBSA OR METRO)	N	Mean	Lower CI	Upper CI
Akron, OH Metropolitan Statistical Area	500	27.30726	26.7878217	27.8267037
Albuquerque, NM Metropolitan Statistical Area	1249	26.93196	26.5882334	27.2756804
Anchorage, AK Metropolitan Statistical Area	491	27.80719	27.2319036	28.3824681
Asheville, NC Metropolitan Statistical Area	765	27.23149	26.7011673	27.7618204
Atlanta-Sandy Springs-Marietta, GA Metropolitan Statistical Area	1792	27.26185	26.9473027	27.5763892
Baltimore-Towson, MD Metropolitan Statistical Area	1565	27.29941	26.9751238	27.6237011
Baton Rouge, LA Metropolitan Statistical Area	706	27.11974	26.6461662	27.5933123
Bethesda-Frederick-Gaithersburg, MD Metropolitan Division	844	26.19875	25.8063242	26.5911759
Birmingham-Hoover, AL Metropolitan Statistical Area	775	27.74568	27.2448666	28.2464971
Boise City-Nampa, ID Metropolitan Statistical Area	1146	27.16517	26.7963232	27.5340088
Boston-Quincy, MA Metropolitan Division	1827	26.67618	26.3528253	26.9995375

MSA Name (CBSA OR METRO)	N	Mean	Lower CI	Upper CI
Bridgeport-Stamford-Norwalk, CT Metropolitan Statistical Area	945	26.38566	26.0587434	26.712578
Burlington-South Burlington, VT Metropolitan Statistical Area	1157	26.7605	26.4328057	27.0881956
Cambridge-Newton-Framingham, MA Metropolitan Division	1343	26.40845	26.1068555	26.7100367
Chambersburg, PA Micropolitan Statistical Area	1244	27.88046	27.4930592	28.2678639
Charleston, WV Metropolitan Statistical Area	551	28.55528	27.934482	29.1760684
Cheyenne, WY Metropolitan Statistical Area	545	27.1796	26.6943481	27.6648435
Chicago-Naperville-Joliet, IL Metropolitan Division	2567	27.24467	26.9963778	27.4929612
Cincinnati-Middletown, OH-KY-IN Metropolitan Statistical Area	815	27.84012	27.2190508	28.4611799
Columbia, SC Metropolitan Statistical Area	482	28.22927	27.3764689	29.0820703
Concord, NH Micropolitan Statistical Area	556	27.12854	26.6477257	27.6093459
Dallas-Plano-Irving, TX Metropolitan Division	814	27.94384	27.4553818	28.4322927
Dayton, OH Metropolitan Statistical Area	550	27.39646	26.7508411	28.0420815
Denver-Aurora, CO Metropolitan Statistical Area	1801	26.70801	26.4437018	26.9723271
Des Moines, IA Metropolitan Statistical Area	599	27.65429	27.1514126	28.1571592
Detroit-Livonia-Dearborn, MI Metropolitan Division	669	28.36508	27.8255315	28.9046246
Dover, DE Metropolitan Statistical Area	1230	27.71752	27.3016727	28.1333753
Durham, NC Metropolitan Statistical Area	591	26.91387	26.0699504	27.75779
Edison, NJ Metropolitan Division	1942	26.98722	26.5711684	27.4032711
Essex County, MA Metropolitan Division	717	27.17317	26.6943068	27.6520324
Fargo, ND-MN Metropolitan Statistical Area	540	27.27568	26.6912782	27.8600813
Florence, SC Metropolitan Statistical Area	812	28.30825	27.8076111	28.8088896
Greensboro-High Point, NC Metropolitan Statistical Area	811	27.3293	26.7974266	27.8611673
Hartford-West Hartford-East Hartford, CT Metropolitan Statistical Area	1918	27.18891	26.8085242	27.5692953
Hilo, HI Micropolitan Statistical Area	1330	27.06617	26.7183236	27.414017
Honolulu, HI Metropolitan Statistical Area	2715	26.70797	26.4554962	26.9604517
Houston-Baytown-Sugar Land, TX Metropolitan Statistical Area	1076	27.59715	27.22026	27.9740337
Huntington-Ashland, WV-KY-OH Metropolitan Statistical Area	547	27.7834	27.2243499	28.3424462
Indianapolis, IN Metropolitan Statistical Area	1208	28.04643	27.6460168	28.4468432
Jackson, MS Metropolitan Statistical Area	658	27.97985	27.3963707	28.5633275
Jacksonville, FL Metropolitan Statistical Area	540	27.74704	27.1236845	28.3703989
Kahului-Wailuku, HI Micropolitan Statistical Area	1173	27.03182	26.6167208	27.4469173
Kansas City, MO-KS Metropolitan Statistical Area	1761	27.70913	27.3573636	28.0608962
Kapaa, HI Micropolitan Statistical Area	575	26.6426	26.1909854	27.094205
Lancaster, PA Metropolitan Statistical Area	1167	27.61778	27.2742836	27.9612719
Las Vegas-Paradise, NV Metropolitan Statistical Area	985	27.52817	27.1212939	27.9350434
Lebanon, NH-VT Micropolitan Statistical Area	982	27.08799	26.5776232	27.5983649
Lincoln, NE Metropolitan Statistical Area	620	27.40985	26.8505727	27.9691288
Little Rock-North Little Rock, AR Metropolitan Statistical Area	847	27.7072	27.2722678	28.1421278
Los Angeles-Long Beach-Glendale, CA Metropolitan Division	951	27.46223	26.9332052	27.9912443
Louisville, KY-IN Metropolitan Statistical Area	855	27.83749	27.3871411	28.2878329
Manchester-Nashua, NH Metropolitan Statistical Area	1315	27.05704	26.7553646	27.3587087
Memphis, TN-MS-AR Metropolitan Statistical Area	785	28.56526	27.75532	29.3751938
Miami-Miami Beach-Kendall, FL Metropolitan Division	669	27.28735	26.7650021	27.8096971

MSA Name (CBSA OR METRO)	N	Mean	Lower CI	Upper CI
Milwaukee-Waukesha-West Allis, WI Metropolitan Statistical Area	1343	27.0996	26.7167658	27.4824311
Minneapolis-St. Paul-Bloomington, MN-WI Metropolitan Statistical Area	2479	27.36621	27.1310104	27.6014149
Nashville-Davidson--Murfreesboro, TN Metropolitan Statistical Area	688	27.203	26.7119189	27.6940758
Newark-Union, NJ-PA Metropolitan Division	1260	27.66259	27.0385165	28.2866631
New Haven-Milford, CT Metropolitan Statistical Area	925	26.62321	26.2473162	26.9991085
New Orleans-Metairie-Kenner, LA Metropolitan Statistical Area	1212	27.66659	27.2693743	28.063797
New York-Wayne-White Plains, NY-NJ Metropolitan Division	3340	26.94076	26.6554735	27.2260388
Norwich-New London, CT Metropolitan Statistical Area	477	27.29212	26.676814	27.9074343
Ogden-Clearfield, UT Metropolitan Statistical Area	515	27.18556	26.5997242	27.7713989
Oklahoma City, OK Metropolitan Statistical Area	1848	27.31148	27.0085279	27.6144381
Omaha-Council Bluffs, NE-IA Metropolitan Statistical Area	1600	27.62069	27.3151845	27.9261893
Orlando, FL Metropolitan Statistical Area	527	26.89466	26.3097228	27.4795953
Philadelphia, PA Metropolitan Division	3031	27.38582	27.0713909	27.7002432
Phoenix-Mesa-Scottsdale, AZ Metropolitan Statistical Area	996	27.21054	26.7840902	27.636988
Pittsburgh, PA Metropolitan Statistical Area	2968	27.84288	27.5255016	28.1602563
Portland-South Portland, ME Metropolitan Statistical Area	785	26.87162	26.477999	27.2652449
Portland-Vancouver-Beaverton, OR-WA Metropolitan Statistical Area	1492	27.28724	26.9186423	27.6558443
Providence-New Bedford-Fall River, RI-MA Metropolitan Statistical Area	4270	27.11925	26.9172659	27.3212259
Rapid City, SD Metropolitan Statistical Area	538	27.00591	26.4349809	27.5768395
Reno-Sparks, NV Metropolitan Statistical Area	996	26.97075	26.6117759	27.329733
Richmond, VA Metropolitan Statistical Area	620	27.98973	27.4986363	28.4808311
Rockingham County-Strafford County, NH Metropolitan Division	1412	26.8153	26.5417266	27.0888633
St. Louis, MO-IL Metropolitan Statistical Area	974	27.81962	27.4212681	28.2179738
Salt Lake City, UT Metropolitan Statistical Area	1500	26.60844	26.2980404	26.9188429
Seaford, DE Micropolitan Statistical Area	1301	28.04717	27.7025929	28.3917363
Seattle-Bellevue-Everett, WA Metropolitan Division	1760	26.80897	26.5227449	27.0952015
Sioux Falls, SD Metropolitan Statistical Area	1050	27.45888	27.1137377	27.804014
Springfield, MA Metropolitan Statistical Area	898	27.22352	26.7611295	27.6859008
Suffolk County-Nassau County, NY Metropolitan Division	544	26.82029	26.3888715	27.2517168
Tacoma, WA Metropolitan Division	480	28.12875	27.4240019	28.8335046
Tampa-St. Petersburg-Clearwater, FL Metropolitan Statistical Area	815	27.00699	26.5712603	27.4427203
Toledo, OH Metropolitan Statistical Area	512	28.1785	27.5454158	28.8115759
Topeka, KS Metropolitan Statistical Area	477	28.26015	27.6513396	28.8689643
Tucson, AZ Metropolitan Statistical Area	695	26.58459	26.1282495	27.0409266
Tulsa, OK Metropolitan Statistical Area	1505	27.4312	27.063778	27.7986186
Virginia Beach-Norfolk-Newport News, VA-NC Metropolitan Statistical Area	878	27.82211	27.2736885	28.370539
Warren-Farmington Hills-Troy, MI Metropolitan Division	887	27.32761	26.9431264	27.7120914

MSA Name (CBSA OR METRO)	N	Mean	Lower CI	Upper CI
Washington-Arlington-Alexandria, DC-VA-MD-WV Metropolitan Division	3747	27.00408	26.5465597	27.4615929
Wichita, KS Metropolitan Statistical Area	797	27.86722	27.4049061	28.3295266
Willimantic, CT Micropolitan Statistical Area	478	27.65834	27.0533107	28.2633667
Wilmington, DE-MD-NJ Metropolitan Division	1388	27.37675	27.0087348	27.7447581
Winston-Salem, NC Metropolitan Statistical Area	485	27.46217	26.6274192	28.2969108
Worcester, MA Metropolitan Statistical Area	989	27.40473	26.9636602	27.8457933
Youngstown-Warren-Boardman, OH-PA Metropolitan Statistical Area	604	27.79112	27.0175198	28.5647209

Because the mean BMI values are based upon survey data, they should always be viewed along with their confidence intervals.

2. Merging BMI Data with Restaurant and Urban Sprawl Data

After calculating the average BMI for each MSA, the BRFSS estimates were merged with the urban sprawl and restaurant data. In other words the restaurant data was matched to their specific MSAs and BRFSS data. Upon merging the datasets it was discovered that only 53 MSAs contained both urban sprawl/restaurant data as well as BMI estimates. This meant that either the MSA was differently defined from the MSA found in the sprawl index and/or the BRFSS data was not available for the specific MSA. Only the 53 MSAs (representing one third of the US population), with the selected data matching the defined criteria were used in the study.

3. Computation of MSA BMI Risk Index

The penultimate step in the analysis involved computation of the BMI Risk Index. Prior to any computation, 80% of the 53 MSA were randomly selected via the surveyselect statement in SAS and were set aside for the BMI Risk Index model development. The other 20% were set aside for model validation of the analysis. The index was obtained by regressing the values of the average MSA BMI on the urban sprawl index for each MSA, the per capita number of limited-service and full-service restaurants for each MSA and their first order interaction. **Table 8** shows the possible independent variables that were eligible for final model inclusion.

Table 8: Independent Variables

Variable	Calculation/Derivation	Description
Sprawl_index	See Appendix1 for details	An indication of Sprawl where larger values indicate less sprawl and smaller values indicate more sprawl
VAR_722211_cap	VAR_722111/population02	The per capital number of Limited-service restaurants defined by the US Census Bureau
VAR_722110_cap	VAR_722110/population02	The per capital number of Full-service restaurants defined by the US Census Bureau
interaction2	(VAR_722110/population02)* Sprawl_index	The interaction of per capita number of full-service restaurants per MSA and Urban Sprawl
Interaction3	(VAR_722211/population02)* Sprawl_index	The interaction of per capita number of limited-service restaurants per MSA and Urban Sprawl

The Adjusted R-squared Technique was used to select the best possible regression model for the data. All possible combinations of the variables listed above and all possible subsets of these variables were included in separate regression models in an iterative fashion. R-squared values (R-squared values are a measurement of the proportion of variation in the data explained by the model) and adjusted R-square values (adjusted R-squared values are similar to R-squared values, only they are adjusted downward to account for the number of variables in the model) were obtained for each model and the models were then rank ordered according to the adjusted R-squared values (see Table 28: SAS Adjusted R-Square). Typically the model with the largest adjusted R-squared value is selected as the final model in this approach. The model with the highest adjusted R-squared value did not include the Sprawl index. However, the second highest-ranked model did include this value and was therefore selected as the final model as it had a higher overall R-squared value and there was no significant drawback in including one additional term in the selected regression model.

The final model selected was represented as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_3 + \beta_5 X_1 X_2$$

Where

Y = The Average BMI for a given Metropolitan Statistical Area;

X_1 = A sprawl index measurement as described in and provided by www.smartgrowthamerica.org;

X_2= The per capita number of limited-service restaurants in a given metropolitan area as provided and described by the US Census Bureau;

X_3 = The per capita number of full-service restaurants in a given metropolitan area as provided and described by the US Census Bureau;

X_4 = The interaction effect obtained by multiplying the per capita number of full-service restaurants in an area by the measure of urban sprawl (X_1 X_3);

X_5 = The interaction effect obtained by multiplying the per capita number of limited-service restaurants in an area by the measure of urban sprawl (X_1 X_2); and

$\beta_1 - \beta_5$ correspond to the coefficients or weights that are applied to each variable.

Table 9: β and P-values

Parameter	Estimate	P-Value
β_0 (Intercept)	27.16074	<.0001
β_1 (Coefficient of sprawl index)	0.00184	0.9035†
β_2 (Coefficient of per capita number of limited-service restaurants)	9616.96792	0.0123
β_3 (Coefficient of per capita number of full-service restaurants)	-8349.99510	0.0083
β_4 (Coefficient of the interaction effect obtained by multiplying the per capita number of full-service restaurants in an area by the measure of urban sprawl)	68.93894	0.0220
β_5 (Coefficient of the interaction effect obtained by multiplying the per capita number of limited-service restaurants in an area by the measure of urban sprawl)	-81.97171	0.0183

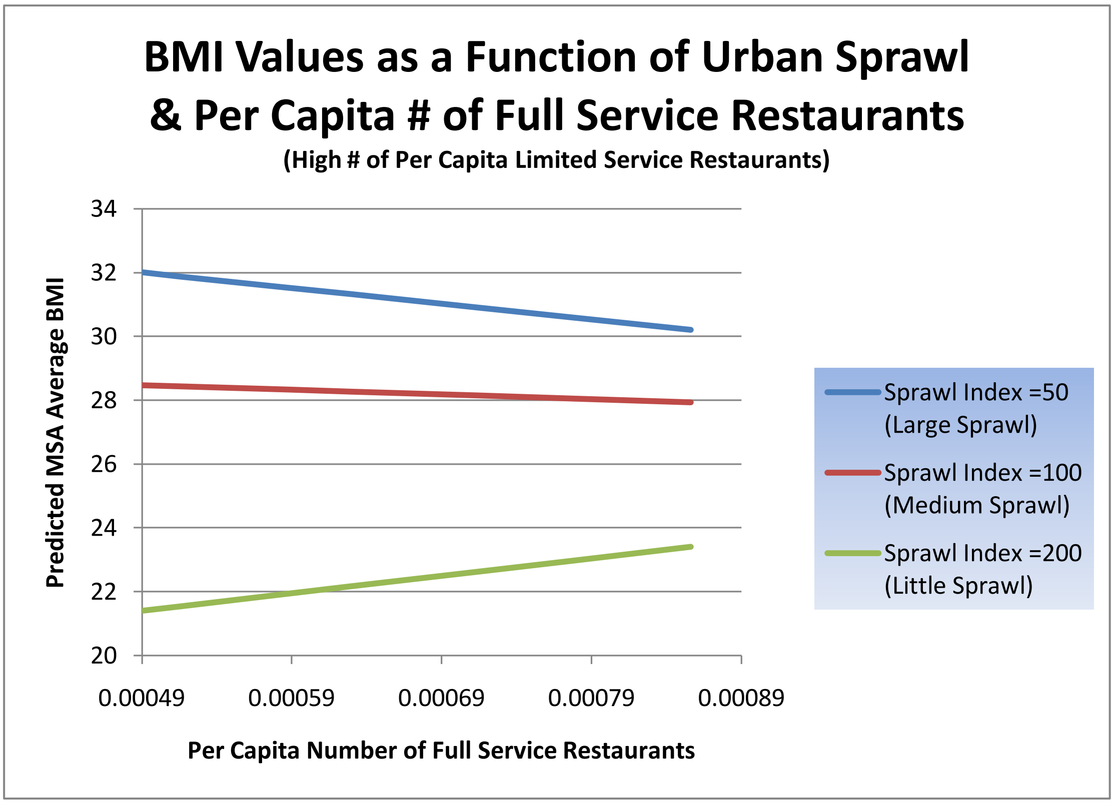
†The p-value for the sprawl index is insignificant at α=0.05, the sprawl-index coefficient is included in the equation for demonstration purposes only, as it has no impact on the outcome.

Values for the model coefficients were obtained and when substituted into the model produced the following equation:

$$Y = 27.16074 + (0.00184)X_1 + (9616.96792)X_2 - (8349.99510)X_3 + (68.93894)X_1X_3 - (81.97171)X_1X_2.$$

To understand this model it helps to view graphs of the model under several different scenarios. **Graph 1** shown below represents the model's predicted MSA Average BMI as a function of the per capita number of full service restaurants for three different Sprawl indices (large, medium, and small sprawl) in an MSA that has a high number of per capita limited-service restaurants (high=0.0013):

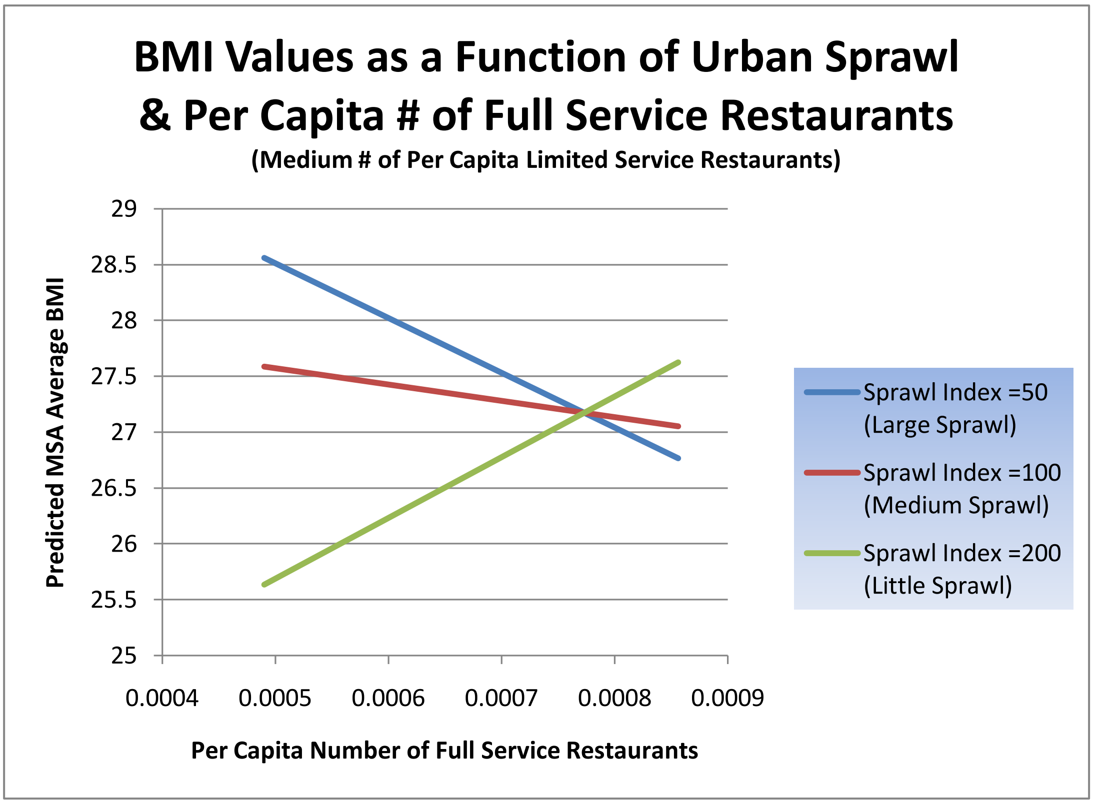
Graph 1: High Number of Per Capita Limited-Service Restaurants



For MSAs with high and medium sprawl values and a high number of limited-service restaurants, as the number of full-service restaurants increases, Ceteris paribus, a decrease in the average community BMI is observed. It should be noted that regardless of the per capita number of restaurants, communities with large sprawl are predicted to have higher average community BMI values than MSAs with less urban sprawl. However, this gap tends to close as more full-service restaurants open and their number increases in the MSA. This same trend can be seen in each graph that we have constructed, suggesting that communities with large urban sprawl tend to have higher values of BMI, when fewer full-service restaurants are found in the community. It appears that the observed effect is reversed in MSAs with little urban sprawl, here we observe that independent of the ratio of limited-service restaurants to the number of full-service restaurants an increase in the number of full-service restaurants results in the rise of the mean MSA BMI.

Graph 2 shows similar information as **Graph 1**, the only difference is that it specifies MSAs with a medium rather than high number of per capita limited service restaurants (medium= 0.00067):

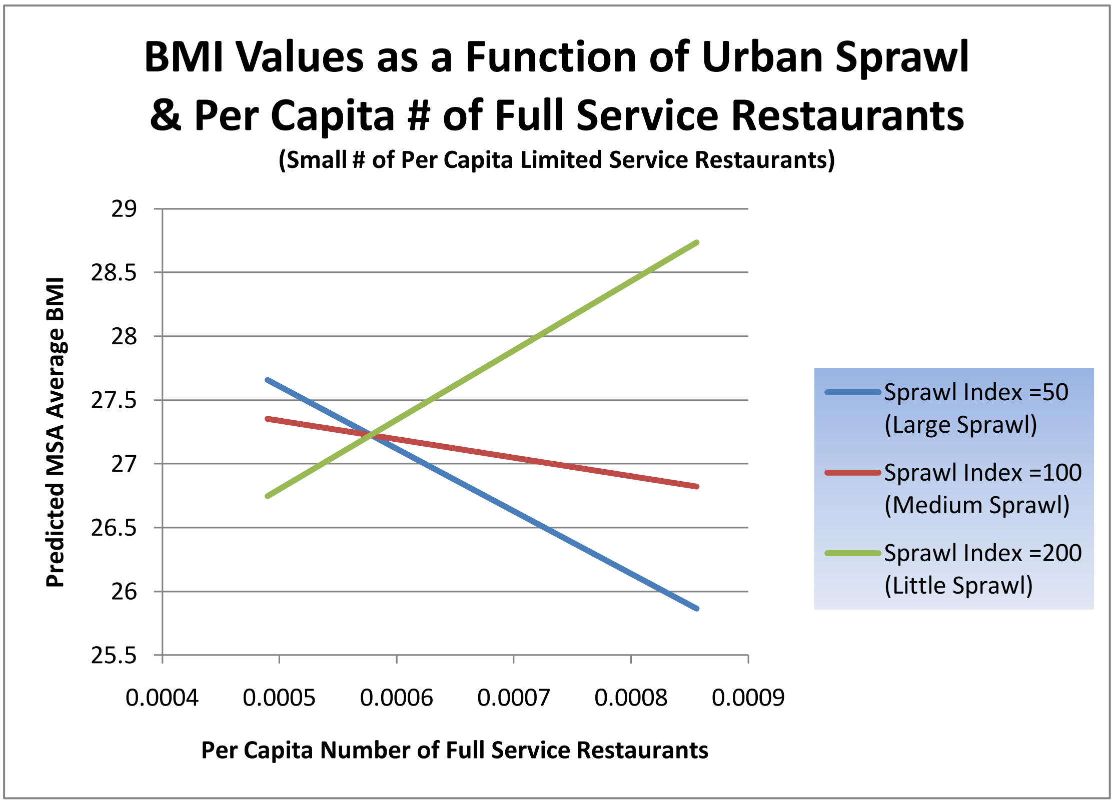
Graph 2: Medium Number of Per Capita Limited-Service Restaurants



It is observed that large sprawl MSAs tend to have larger predicted community mean BMIs than medium and little sprawl MSAs. However, these gaps close and eventually reverse their trend as the number of full-service restaurants increases.

Graph 3 shows the effect of increasing the number of full-service restaurants in a community where there are few limited-service restaurants (low=0.00051). A similar trend as in the preceding graphs is observed. In this scenario when observing MSAs with little sprawl, increasing the number of full-service restaurants increases the overall BMI of the community rather drastically.

Graph 3: Small Number of Per Capita Limited-Service Restaurants



An attempt was made to fit a reduced model that examined the Sprawl Index, the per capita number of *total* restaurants [(full-service restaurants + limited-service restaurants)/population - without breaking out the restaurants by full-service and limited-service restaurants], and the interaction between the per capita number of *total* restaurants and the Sprawl index. When doing so, no terms (with the exception of the intercept term, as would be expected) were found to be statistically significant and the overall model failed to reach any level of statistical significance.

On further detailed analysis it was discovered that on average a 0.0001 per capita increase in the number of full-service restaurants in a community with high/medium/low per capita number of limited-service restaurants, Ceteris paribus, results in an average decrease in BMI of 0.46 kg/m^2 for high sprawl areas, a decrease of 0.14 kg/m^2 for medium sprawl communities and an increase of 0.52 kg/m^2 for low sprawl communities.

4. Correlation Analysis – Pearson Correlation Coefficients

Table 10 shows a list of simple summary statistics for each variable included in the final BMI Index model. The table presents figures such as the means, standard deviations, minimum and maximum values as well as the variable descriptions.

Table 10: Simple Statistics

6 Variables:	Mean Spawl_index VAR_722211_cap VAR_722110_cap interaction2 interaction3						
Simple Statistics							
Variable	**N**	**Mean**	**Std Dev**	**Sum**	**Minimum**	**Maximum**	**Label**
Mean	43	27.34920	0.53437	1176	26.38566	28.56526	Average BMI
Spawl_index	43	102.58512	21.01270	4411	46.78000	140.21000	Sprawl Index
VAR_722211_cap	43	0.0006962	0.0001240	0.02994	0.0005177	0.00130	Per Cap - Limited-service Restaurants
VAR_722110_cap	43	0.0006447	0.0001869	0.02772	0	0.00147	Per Cap - Full-service Restaurants
interaction2	43	0.06517	0.02016	2.80245	0	0.10841	interaction of per capita number of full-service restaurants per MSA and Urban Sprawl
interaction3	43	0.07081	0.01731	3.04490	0.03854	0.13269	interaction of per capita number of limited-service restaurants per MSA and Urban Sprawl

The output of **Table 11** shows the correlation analyses. In this table one sees a list of all the variables in the model across the top of the table and the same list duplicated on the side of the table. When examining the intersection between any variable on the x-axis of this table (columns) and any other variable on the y-axis (row), we find a cell which lists the Pearson Correlation Coefficient at the top. The Pearson Correlation Coefficient is a figure between -1 and +1 and measures the linear association or relationship between two variables. A value of +1 or -1 indicates a perfect linear relationship and a value of zero indicates no linear relationship. The “trace” or diagonal values on the table are the intersection of each variable with itself; the values will always be one, since a value must be perfectly correlated with itself. Underneath each correlation coefficient one finds the p-value associated with a statistical test which determines whether there is evidence that there is a significant linear relationship between two variables; i.e. if there is evidence that the two variables’ linear correlation is statistically significant (different from zero). Small values of the p-value show evidence that there is strong linear relationship, regardless of whether the relationship is positive or negative (i.e. one value increases with the other or one value increases while the other decreases).

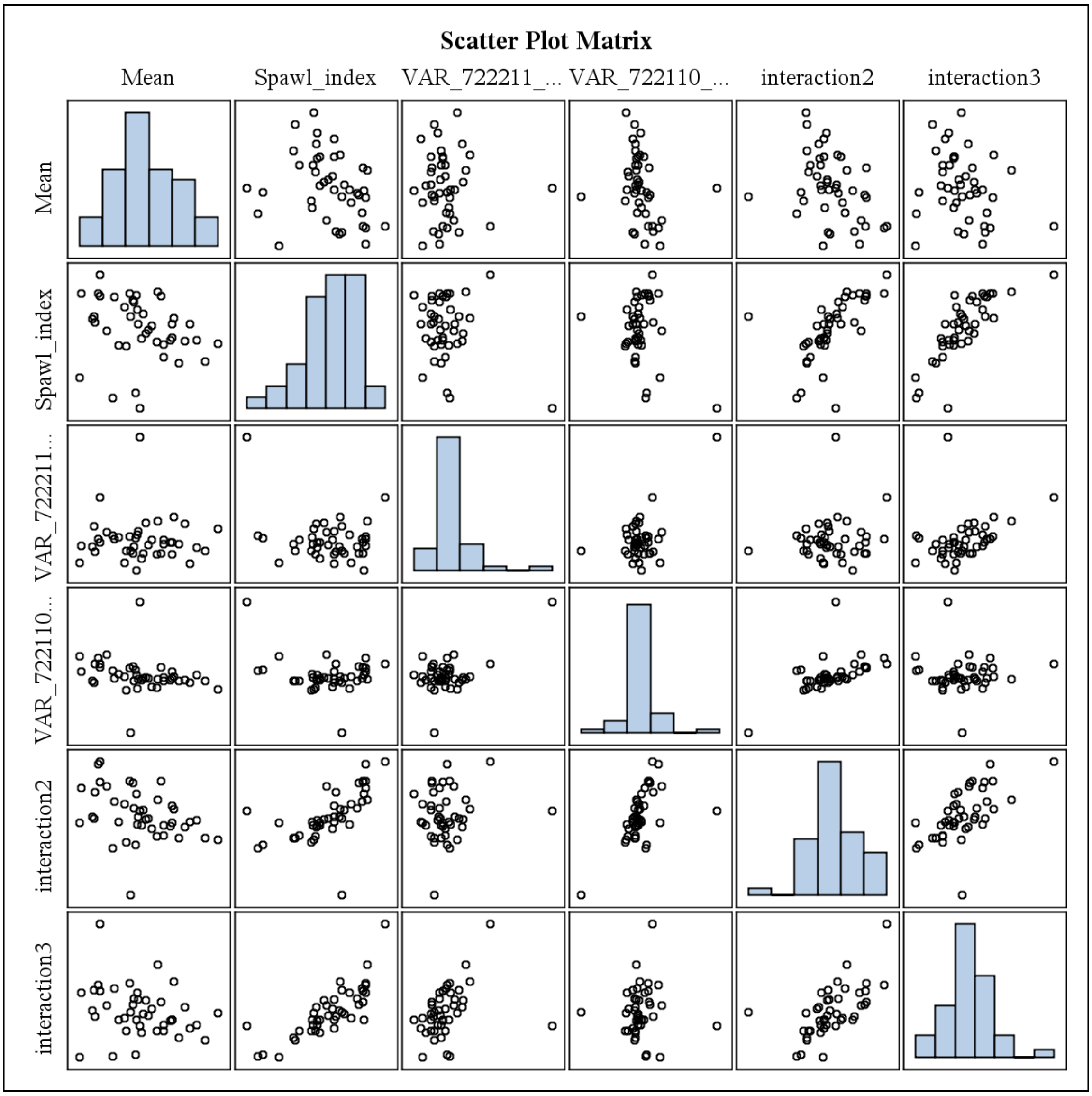
As an example of how to read this output, one can examine row VAR_722211_cap (Per Cap - Limited-service Restaurants) and column Spawl_index (Sprawl Index) in **Table 11**. The table cell which intersects this row and column shows a value of -0.25152 which is the Pearson Correlation Coefficient. The p-value is 0.1037, which is typically considered to be a large value. As a result, we find that there is little evidence of a strong linear relationship between the per capita number of limited-service restaurants and the urban sprawl index. As expected there seems to be some fairly large associations between the interaction terms (e.g. interaction2) and the first order variables that make up the interaction terms [e.g. interaction of per capita number of full-service restaurants per MSA (VAR_722110_cap) and Urban Sprawl (Urban Sprawl)].

Table 11: Correlation Analyses

Pearson Correlation Coefficients, N = 43 Prob > |r| under H0: Rho=0						
	Mean	**Spawl_index**	**VAR_722211_cap**	**VAR_722110_cap**	**interaction2**	**interaction3**
Mean Average BMI	1.00000	-0.23894 0.1228	-0.01625 0.9176	-0.24231 0.1175	-0.37425 0.0134	-0.22972 0.1384
Spawl_index Sprawl Index	-0.23894 0.1228	1.00000	-0.23815 0.1241	-0.25152 0.1037	0.61619 <.0001	0.79715 <.0001
VAR_722211_cap Per Cap - Limited-service Restaurants	-0.01625 0.9176	-0.23815 0.1241	1.00000	0.55822 0.0001	0.12388 0.4287	0.34762 0.0224
VAR_722110_cap Per Cap - Full-service Restaurants	-0.24231 0.1175	-0.25152 0.1037	0.55822 0.0001	1.00000	0.56486 <.0001	0.00684 0.9653
interaction2 interaction of per capita number of full-service restaurants per MSA and Urban Sprawl	-0.37425 0.0134	0.61619 <.0001	0.12388 0.4287	0.56486 <.0001	1.00000	0.61983 <.0001
interaction3 interaction of per capita number of limited-service restaurants per MSA and Urban Sprawl	-0.22972 0.1384	0.79715 <.0001	0.34762 0.0224	0.00684 0.9653	0.61983 <.0001	1.00000

The final piece of output is the correlation matrix. This is displayed in **Table 12** and shows the same information as **Table 11** only in graphical form. Each row and column intersection shows a scatter plot of the variables listed on the x (column) and y axis (row). We observe that the variables interaction2 and Sprawl-index are more highly correlated than the Sprawl_Index and VAR_722211_cap.

Table 12: Scatter Plot Matrix



Studying the various fields gives an idea of how the variables are correlated.

The following statements can be made;

- There is a low degree of correlation between the mean BMI and all the other five variables.
- There is a low degree of correlation between the Sprawl_Index and per cap full-service restaurants and per cap limited-service restaurants, but a high degree of correlation between interaction2 and interaction3, which is to be expected.
- There is a relatively high degree of positive correlation between per cap limited-service restaurants and per cap full-service restaurants, a low correlation between per cap limited-service restaurants and interaction2 and a medium degree of correlation between per cap limited-service restaurants and interaction3.
- There is a high degree of correlation between per cap full-service restaurants and interaction2 and a low degree of correlation between per cap full-service restaurants and interaction3.

The degree of correlation that needs further explanation is the high degree of positive correlation between per cap full-service restaurants and per cap limited service restaurants. A plausible explanation could be that full-service restaurants are located in the same area as limited-service restaurants; this is obvious due to the commercial zoning laws in communities predetermining these locations. Furthermore, as demand for limited-service restaurants increases so does the demand for full-service restaurants and vice versa, as has been the case from 1972 to 1997 (Bureau of the Census, 1976, 2000) when the number limited-service restaurants doubled and the number of full-service restaurants increased by 35%.

5. Model Validation

In the final analysis of this study, it is examined how well this model could predict data that was not used in its development. For this purpose, the model was applied to the 20% holdout sample discussed at the beginning. Predictions of the community BMI were made from the model for each MSA in the holdout sample. These predicted values were recorded and then examined to determine if they fell within the 95% confidence interval of the actual community BMI estimated from the BRFSS data. All but one of the 10 predictions i.e., 90% fell within the 95% confidence interval, suggesting the model fits the data fairly well. The predicted values and the upper and lower confidence intervals for the predictions are shown in the table below. The incorrectly predicted MSA is highlighted in red.

Table 13: Validation

Metropolitan Statistical Area	Lower CI	Upper CI	Predicted Value
Birmingham-Hoover, AL MSA	27.2448666	28.2464971	27.8368958
Jacksonville, FL MSA	27.1236845	28.3703989	27.3936138
Orlando, FL MSA	26.3097228	27.4795953	27.2707876
Phoenix-Mesa-Scottsdale, AZ MSA	26.7840902	27.6369880	27.3287743
Portland-Vancouver-Beaverton, OR-WA MSA	26.9186423	27.6558443	27.1671958
Tampa-St. Petersburg-Clearwater, FL MSA	26.5712603	27.4427203	26.9991764
Toledo, OH MSA	27.5454158	28.8115759	27.2191400
Virginia Beach-Norfolk-Newport News, VA-NC MSA	27.2736885	28.3705390	27.3432033
Winston-Salem, NC MSA	26.6274192	28.2969108	27.3657534
Worcester, MA MSA	26.9636602	27.8457933	27.2916659

6. Post Study Processing: BMI Risk Index

At the conclusion of the analysis, the BMI risk model was applied to all MSAs in the study. The resulting values were then transformed to a 0 to 100 point scale, using the following formula:

$$\text{BMI Risk Index Score for MSA } Y = \frac{Y \text{ Predicted BMI Score} - min_{(\text{all } Y)}(\text{Predicted BMI Score})}{max_{(\text{all } Y)}(\text{Predicted BMI Score}) - min_{(\text{all } Y)}(\text{Predicted BMI Score})}$$

This formula simply sets the MSA with the largest Predicted Risk to 100, the one with the smallest risk to 0 and scales all other MSAs accordingly. When applied to the 53 MSAs in this study, Bridgeport-Stamford-Norwalk CT is ranked the least risky and Memphis, TN becomes the most risky. After Memphis, Birmingham, AL is the second most risky in this study*.

* *It should be noted that the findings that Memphis and Birmingham are the most risky cities is consistent with a 2007 Forbes study which found them to be the fattest cities in America (http://www.forbes.com/2007/11/14/health-obesity-cities-forbeslife-cx_rr_1114obese.html).*

The table below shows all MSAs used in the analysis and their corresponding BMI Risk Index, sorted from least risk to highest risk.

Table 14: BMI Risk Index

Rank	MSA Name	BMI Risk Index
1	Bridgeport-Stamford-Norwalk, CT Metropolitan Statistical Area	0.00
2	Honolulu, HI Metropolitan Statistical Area	15.16
3	Omaha-Council Bluffs, NE-IA Metropolitan Statistical Area	35.81
4	Tampa-St. Petersburg-Clearwater, FL Metropolitan Statistical Area	38.04
5	Albuquerque, NM	46.12
6	Essex County, MA Metropolitan Division	46.25
7	Seattle-Bellevue-Everett, WA Metropolitan Division	46.38
8	New Haven-Milford, CT Metropolitan Statistical Area	46.92
9	Cambridge-Newton-Framingham, MA Metropolitan Division	48.06
10	Denver-Aurora, CO Metropolitan Statistical Area	48.63
11	Portland-Vancouver-Beaverton, OR-WA Metropolitan Statistical Area	49.37
12	Boston-Quincy, MA Metropolitan Division	49.91
13	Chicago-Naperville-Joliet, IL Metropolitan Division	50.01
14	Springfield, MA Metropolitan Statistical Area	51.14
15	New Orleans-Metairie-Kenner, LA Metropolitan Statistical Area	51.60
16	Pittsburgh, PA Metropolitan Statistical Area	52.26
17	Milwaukee-Waukesha-West Allis, WI Metropolitan Statistical Area	52.32
18	Greensboro--Winston-Salem--High Point, NC Combined Statistical Area	52.76
19	Toledo, OH Metropolitan Statistical Area	52.87
20	Miami-Miami Beach-Kendall, FL Metropolitan Division	54.40
21	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metropolitan Statistical Area	54.57
22	Orlando, FL Metropolitan Statistical Area	56.35
23	Baltimore-Towson, MD Metropolitan Statistical Area	57.00
24	Tacoma, WA Metropolitan Division	57.68
25	Worcester, MA Metropolitan Statistical Area	57.76
26	Tucson, AZ Metropolitan Statistical Area	59.64
27	Phoenix-Mesa-Scottsdale, AZ Metropolitan Statistical Area	60.26
28	Minneapolis-St. Paul-Bloomington, MN-WI Metropolitan Statistical Area	60.72
29	Virginia Beach-Norfolk-Newport News, VA-NC Metropolitan Statistical Area	61.23
30	Columbia, SC Metropolitan Statistical Area	62.70
31	Winston-Salem, NC Metropolitan Statistical Area	62.75
32	Salt Lake City, UT Metropolitan Statistical Area	63.31
33	Wichita, KS Metropolitan Statistical Area	63.32

Rank	MSA Name	BMI Risk Index
34	Los Angeles-Long Beach-Glendale, CA Metropolitan Division	64.20
35	Jacksonville, FL Metropolitan Statistical Area	64.63
36	Akron, OH Metropolitan Statistical Area	66.97
37	Cincinnati-Middletown, OH-KY-IN Metropolitan Statistical Area	69.13
38	Las Vegas-Paradise, NV Metropolitan Statistical Area	71.00
39	Kansas City, MO-KS Metropolitan Statistical Area	71.11
40	Washington-Arlington-Alexandria, DC-VA-MD-WV Metropolitan Statistical Area	71.88
41	St. Louis, MO-IL Metropolitan Statistical Area	73.10
42	Houston-Baytown-Sugar Land, TX Metropolitan Statistical Area	75.90
43	Tulsa, OK Metropolitan Statistical Area	77.31
44	Atlanta-Sandy Springs-Marietta, GA Metropolitan Statistical Area	79.01
45	Detroit-Warren-Livonia, MI Metropolitan Statistical Area	79.17
46	Ogden-Clearfield, UT Metropolitan Statistical Area	85.08
47	Baton Rouge, LA Metropolitan Statistical Area	86.21
48	Dallas-Plano-Irving, TX Metropolitan Division	86.58
49	Little Rock-North Little Rock, AR Metropolitan Statistical	87.96
50	Indianapolis, IN Metropolitan Statistical Area	88.43
51	Durham, NC Metropolitan Statistical Area	90.62
52	Birmingham-Hoover, AL Metropolitan Statistical Area	94.51
53	Memphis, TN-MS-AR Metropolitan Statistical Area	100.00

7. Discussion, Limitations and Conclusion

Discussion

The study was to determine whether a BMI risk index could be created based on variables that were established analyzing prior research. It was shown that a robust correlation between the proxy-variables used and the mean BMI of the MSA population exists and that the developed index has predictive power. The results of this study contribute to the notion that there is a positive association between limited-service restaurants and rise in BMI as well as a positive association between high suburban sprawl and high BMI as indicated in prior cross-sectional research papers (Chang & Mehta, 2008, Chou, 2004, Eid, 2007, Ewing R. B., 2006, Ewing R. S., 2003, Jeffery, 1998, Kyureghian, 2007, Lakdawalla D. P., 2005, Lopez, 2004, Philipson, 1999, Giles-Corti, 2003, Saelens, 2003, Frank, 2004). The study has taken a further step towards identifying and using proxy-variables in a model that gives us an enhanced view on the possible determining factors that underlie the rise in BMI in MSAs and enables ranking of MSAs according to those intrinsic risks. The results show that we have to differentiate between full-service and limited-service restaurants when evaluating their impact on rise in BMI. Independent of the level of urban sprawl, limited-service restaurants have a higher positive correlation to an increased community BMI than full-service restaurants. This study goes further to note that full-service restaurants have a negative correlation to the mean community BMI of populations that reside in large to medium degrees of urban sprawl, but have a positive correlation to the mean community BMI of a population with little degree of urban sprawl – an observation that is not entirely intuitive. As Chang et al (2008) mentioned in their study, more research has to be done in defining the nutritional differences between food served in limited-service restaurants and full-service restaurants. Similar to the findings of Chang et al (2008) this study shows that the ratio of the number of full-service restaurants to limited-service restaurants is of relevance. By being more granular in the analysis of geographical locations through the introduction of levels of urban sprawl, this study goes beyond prior research in this area. It also shows the complex relationships between factors that influence the rise in overweight and obesity in a population. The reason why an increase in full-service restaurants has the opposite effect on MSAs with little sprawl as compared to MSAs with high and medium sprawl, independent of the density of limited-service restaurants, requires more research. Whether it is linked to commercial zoning which may favor one of the restaurant types, or whether it is linked to the social makeup of the community and their preferences has yet to be studied. The research shows that when using aggregate restaurant data (sprawl index, per capita total number of restaurants) no terms were found to be statistical significant, contrary to research by Chang et al. (2008) who reported a negative association to rise in BMI and Chou et al. (2004) who indicated a positive association to rise in BMI. These results suggest that when modeling community BMI values, researchers should precede with caution when using aggregated restaurant data as the effects of per capita number of limited service restaurant types and full service restaurant types might cancel one another out when combined.

Research on obesity and overweight is being done in other Organization for Economic Cooperation and Development (OECD) countries where obesity has grown over the past two decades by 8% (Bleich 2007). In our preview we found that most of the research in this area was done in the United States of America. The possible reason for this could be that the United States are leading the world in this epidemic, having the greatest proportion of overweight and obese in its population when compared to other countries. Germany is leading Europe in having the largest proportion of overweight in the population with the United Kingdom leading Europe with the largest proportion of obese (Brunello G. 2008). Utilizing prior research done in the United States, German researchers could exploit cross-sectional data collected through the census bureau and the Gesetzlichekrankenkassen (Health Insurance Companies) to study overweight and obesity in the German population. The fact that Germany has chronological continuous, digitalized data on its population, held by health insurance companies and the census bureau would enable the quality of research that could evolve to be much more defined. Another fact is that urbanization in Germany has developed in a more controlled and curtailed fashion than in the United States, this could bring new insights on the influence of urban sprawl and the built-environment to the rise in the populations weight status.

The findings of this research could help public health organizations better understand the interactions of underlying factors that are part of the cause of the rapid rise in overweight and obesity. By further studying MSAs that have a better ranking, based on the predetermined proxy-variables, changes could be implemented to improve lower performing MSAs. Our findings also show that the public perception that limited-service restaurants are one of the major contributors to BMI appear to be justified – more so than full-service restaurants.

The increase in overweight and obesity is the joint effect of individual behavior and external environment (Chou 2004). In the case of MSAs the effects were induced by changes that evolved over the past 30 years, associated with the chronological development of suburban communities.

The great move of the “middle-class” to Suburbia was initiated by the GI-Bill in the late 1940s, when prefabricated housing was built for WWII veterans and various subsidies distributed to spur economic growth in the U.S. There was a marked change in suburban housing from the 1980s-1990s, a time period where the BMI of the population exploded from 24.94kg/m^2 to 27.07kg/m^2 and obesity rose from 11.05% to 24.04% (Chou 2004). The number of limited-service restaurants doubled around this period (1972-1997) as did the character of suburbia. Townhouses and apartment complexes started to dominate the landscape as more people moved to suburbia. Workers increasingly had to commute to work by car a trend that has increased over time and has been sustained by the relative cheap price of gasoline – in the U.S. 1990 Census the average time to commute to work for the 50 largest MSAs by population was 23.2 minutes. Real income declined for a single worker – the U.S. Census Bureau estimated that real income was 1% higher in 1999 than 1970, inducing increased participation of women in the workforce. These combined effects, caused by economic pressures, make home time a relatively scarce resource and generate the type of substitution effects towards convenience and fast food that accompany an increase in hours of work and commuting. The sprawl of malls with limited service restaurants in high time value suburban communities is indicative of the resulting demand, as is the accompanied higher BMI in those populations, as shown in our study.

To change the status quo will be a great challenge to public health organizations and policymakers. Urban sprawl and its multitude of negative impacts on people and the environment has been addressed by many organizations such as Smart Growth America, but suburban living is an accepted norm in the U.S. culture, as are long commuting times. Changing social norms has always been challenging and takes a generation to achieve, inducing changes in an industry adds to the challenge. For many people eating-out at limited-service restaurants has become more of a necessity than a choice, due to the increased value of time and economic pressures. Entrepreneurs establish restaurant businesses to make a profit and will do so by supplying meals that are in demand at the right price with an optimal profit margin. In recent years State and Federal regulators were able to ban trans-fats (trans-isomer fatty acids) from being used in restaurants in the U.S. Trans-fats are a cheap substitute for more expensive natural fats with the advantage of enhancing taste and having a longer shelf life, but on the downside are grossly unhealthy. It took regulators years to battle the industry and push through legislation to protect consumers. These are the real challenges facing public health organizations and policymakers. If limited-service restaurants were incentivized to serve healthier foods at the right price, under the same time saving concept, positive changes could be made to the benefit of public health and interest. In the meantime further studies are required to explore models that could help motivate change in individual behavior, as this is the area which holds the most promise for effective intervention, but will prove to be the most challenging.

Limitations

There are limitations that have to be considered.

- The model used in deriving the index should not be regarded as indicating causality, as the factors are contributing to Y (Y=The mean BMI for a given Metropolitan Statistical Area) but should not be considered the cause of Y.
- The mathematical equation applied is not necessarily the best representation, but is utilized because;
    - The statistical theory is well developed
    - It functions well and is useful in practice
    - It shows a simple way in which Y relates to a range of different factors
- The individual- and higher-level data used in the models reviewed is derived from various years and collected from several population surveys and censuses that have their individual limitations. For example the Behavior Risk Factor Surveillance Service (BRFSS) has the following limitations;
    - BRFSS generally oversamples less populous states,

    - Geographically identifiers at a finer level than the county are not available,
    - Self-reported anthropometric variables are underreported and have to be adjusted using actual physical measurements from NHANES.
- The index will be developed for predicting the risk of increased body mass in a metropolitan statistical area, this is a subset of the national population and structurally divided from “other populations”, e.g. in rural areas.
- The assumption is made that the variables on the right hand side of the equation are measured without error and the data are “correct” to a reasonable degree – this is never true with real data especially in sample survey data.
- A cross section analysis over time i.e. longitudinal study would have been helpful in enhancing the validity of this study, this was not possible as the data for the sprawl index was only available for the year 2002 and the complexity of recalculating the MSI for other years was beyond the scope of this study.
- This is a cross-sectional study indicating that the association seen between the independent variables (per capita number of full-service, limited-service restaurants, urban sprawl) and the mean BMI cannot be attributed to causality.
- There is a degree of ecological bias due to the fact that inferences made from our model about individuals living in the chosen MSAs can be misleading. This bias, also referred to as ecological fallacy results from the census data being in aggregate form which can misinterpret the underlying relationship between our chosen covariates and the rise in overweight and obesity, represented by increased BMI (Piantadosi S 1988). This could be caused by a small subgroup of extremely obese in the population of an MSA being responsible for a large proportion of the increase in the mean BMI value within that particular area. The ecological fallacy could cause inflated parameter estimates or

parameter estimates that suggest a relationship that is counter intuitive to known relationships.

- We are aware of false inferences that could be introduced by extraneous and confounding variables that may not be known to us and were therefore not controlled for. We believe that we have reduced the systematic error by using the BRFSS survey, which is a randomized survey, and using for the statistical analysis a relatively large sample population of 111,704 people.

Conclusion

The research conveys new insights to the complexity of the association between urban sprawl and per capita full-service and limited-service restaurants and their impact on the BMI in MSAs. Furthermore, the model does predict the community BMI of the MSAs fairly well. This study found that increase in urban sprawl is positively correlated to a rise of the mean community BMI. Furthermore, restaurant mix becomes a positive or negative determinant of the communities BMI depending on the ratio of full-service restaurants to limited-service restaurants; this gets compounded by the degree of urban sprawl found in the MSAs. In high to medium level of urban sprawl a high ratio of limited-service restaurants to full-service restaurants is associated with a rise in BMI, as the number of full-service restaurants increase in the community - keeping the number of limited-service restaurants fixed - the mean BMI of the MSA sinks indicating a negative correlation. On the other hand, for MSAs with little urban sprawl, independent of the density of limited-service restaurants, increasing the number of full-service restaurants is positively correlated to the rise of the community BMI. Further research has to be done to investigate the socioeconomic and environmental factors that might influence eating behaviors and restaurant preferences within different communities and their effect on the population’s weight status.

V. **Works Cited**

Amarasinghe A, D'Souza G,Brown C, Oh H. The Influence of Socioeconomic and Environmental Factors on Health and Obesity in Rural Appalachia. West Virginia University 2006.

Anderson PM, Butcher KF, Levine PB. Maternal Employment and Overweight Children. NBER Work Pap No. 8770, 2002.

Baum II CL, Ruhm CJ. Age, Socioeconomic Status and Obesity Growth. NBER Work Pap No. 13289, 2007.

Bednarek HL, Jeitschko TD, Pecchenino RA. Gluttony and Sloth:signs of trouble or evidence of bliss? A Theory of Choice in the Presence of Behavioral Adjustment Costs. BEJ Econom Anal & Policy, (Contributions to Economic Analysis & Policy, Vol.6) 2006;5(1):1507-1507.

Bleich, S, Cutler, D, Murray, C, Adams, A. Why is the Developed World Obese? NBER Work Pap No. 12954, 2007.

Brunello G., Michaud P-C, Sanz-de-Galdeano A. The Rise in Obesity across the Atlantic: An Economic Perspective. IZA DP No. 3529, 2008.

Bureau of the Census, US Department of Commerce. Census of Retail Trade. Washington, DC: US Government Printing Office, 2002.

Cawley, J, Markowitz, S, Tauras, J. Lighting Up and Slimming Down: The Effects of Body Weight and Cigarette Prices on Adolescent Smoking Initiation. NBER Work Pap No. 9561, 2003.

Chang VW, Mehta NK. Weight Status and Restaurant Availability: A Multilevel Analysis. Am J Prev Med 2008;34(2):127-133.

Chou S, Grossman M, Saffer H. An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System. J Health Econo 2004; 23:565-587.

Conley D, Glauber G. Gender, Body Mass and Economic Status. NBER Work Pap No. 11343, 2005.

Cutler D, Glaeser E, Shapiro J. Why have Americans become more Obese? NBER, Work Pap No. 9446, 2003.

Dietz WH, Gortmaker SL. Do we fatten our children at the television set? Obesity and television viewing in children and adolescent. Pediatrics 1986;75:807-812.

Doh Y, Nayga RM. A Note on The Effect of Television Viewing and Computer Use on Adult Obesity. Texas A & M University College Station 2007.

Dunn R A. Obesity and the Availability of Fast-food:an instrumental variables approach. University of Wisconsin-Madison 2007.

Eid J, Overman HG, Puga D, Turner MA. Fat City: Questioning the Relationship Between Urban Sprawl and Obesity. University of Toronto 2007.

Everitt BS, Dunn G. Applied Multivariate Data Analysis. Arnold, 2001.

Ewing R, Brownson RC, Berrigan D. Relationship Between Urban Sprawl and Weight of United States Youth. Am J Prev Med 2006;31(6): 464-474.

Ewing R, Schmid T, Killingsworth R, Zlot A, Raudenbush S. Relationship between urban sprawl and physical activity, obesity, and morbidity. Am J Health Promot 2003;18:47-57.

Ewing R, Pendall R, Chen D. Measuring Sprawl and its Impact. Smart Growth America 2002.

Finkelstein EA, Fiebelkorn IC, Wang G. National medical spending attributable to overweight and obesity: How much, and who's paying? Health Aff 2003;W3:219-226.

Flegal KM. The Effects of Changes in Smoking Prevalence on Obesity Prevalence in the United States. Am J Public Health 2007;97(8):1510-1514.

Frank LD, Andresen MA, Schmid TL. Obesity relationships with community design, physical activity, and time spent in cars. Am J Prev Med 2004;27(2):87-96.

Gelbach JB, Klick J, Stratmann T. Cheap Donuts and Expensive Broccoli: The Effect of Relative Prices on Obesit. Available at SSRN: http://ssrn.com/abstract=976484, 2007.

Gibson D. Food Stamp Program Participation is Positively Related to Obesity in Low Income Women. School of Public Affairs, Baruch College, City University of New York 2003.

Giles-Corti B, Donovan R J. Relative Influences of Individual, Social Environmental, and Physical Environmental Correlates of Walking. Am J Public Health 2003;93(9):1583-1589.

Gruber J, Frakes M. Does Falling Smoking Lead to Rising Obesity? NBER Work Pap No. 11483, 2005.

Guthrie J, Biing-Hwan Lin, Frazao E. Role of Food Prepared Away from Home in the American Diet, 1977-78 versus 1994-96: Changes and Consequences. J Nutr Educ Behav 2002;34:140-150.

Haines PS, Hungerford DW, Popkin BM, Guilkey DK. Eating Patterns and Energy and Nutrient Intakes of US Women. J Am Diet Assoc 1992;698.

Handy SL, Boarnet MG, Ewing R, Killingsworth RE. How the built environment affects physical activity: views from urban planning. Am J Prev Med 2002;22:88-99.

Hill JO, Wyatt,HR, George W, Reed GW, Peters JC. Obesity and the Environment: Where Do We Go from Here? Science Magazine 2003;299(5608):853-855.

International Obesity Task Force. EU Platform on Diet, Physical Activity and Health. International Obesity Task Force 2005.

Jeffery RW, French SA. Epidemic obesity in the United States: are fast foods and television viewing contributing? Am J Public Health 1998;88(2):277-280.

Kaur H, Choi W, Mayo M, Harris K. Duration of Television Watching is Associated with Increased Body Mass Index. J Pediatr 2003;143(4):506-511.

Kaushal N. Do food stamps cause obesity? Evdence from immigrant experience. NBER, Work Pap No 12849, 2007.

Kuchler F, Golan E. Is There a Role for Government in Reducing the Prevalence of Overweight and Obesity? Choices, an online magazine of the American Agricultural Economics Association 2004;19.

Kyureghian G, Nayga RM, Davis GC, Biing-Hwan Lin. Food Away from Home Consumption and Obesity: An Analysis by Service Type and by Meal Occasion. American Agricultural Economics Association, Annual Meeting 2007.

Labor, U.S. Department of. Employment, Hours, and Earnings from the Current Employment Statistics survey. National. Bureau of Labor Statistics 1970-2003.

Lakdawalla D, Philipson T. The Growth of Obesity and Technological Change: A Theoretical and Empirical Examination. NBER Work Pap No. 8946, 2002.

Lakdawalla D, Philipson T, Bhattacharya J. Welfare-Enhancing Technological Change and the Growth of Obesity. Rand Corporation, www.rand.org, 2005.

Lopez, R. Urban Sprawl and Risk for Being Overweight or Obese. Am J Public Health 2004;94:1574-1579.

Loureiro ML, Nayga RM. International Dimensions of Obesity and Overweight Related Problems: An Economics Perspective . Am J Agric Econ 2005;87(5):1147-1153.

Mandal B, Chern WS. Econometric Analysis of Rising Body Mass Index in the U.S.: 1996 versus 2002. Consumer Interests Annual 2006;51.

Mandal B, Chern WS. A Multilevel Approach to Model Obesity and Overweight in the United States. Available at SSRN: http://ssrn.com/abstract=989227 2007.

Morland K, Wing S, Diez Roux A, Poole C. Neighborhood characteristics associated with the location of food stores and food service places. Am J Prev Med 2002;22:23-29.

Papas MA, Alberg AJ, Ewing R, Helzlsouer KJ, Gary TL, Klassen AC. The built environment and obesity. Epidemiol Rev 2007;29:129-43.

Philipson TJ, Posner RA. The long-run growth in obesity as a function of technological change. NBER Work Pap No. 7423, 1999.

Piantadosi S, Byar DP, Green SB. The Ecological Fallacy. Am J Epidemiol 1988;127:893-904.

Powell LM, Auld MC, Chaloupka FJ, O'Malley PM,Johnston LD. Access to Fast Food and Food Prices: Relationship with Fruit and Vegetable Consumption and Overweight among Adolescents. Adv Alcohol Subst Abuse 2006;17:23-48.

Proctor MH, Moor LL, Gao D, Cupples LA. Television viewing and change in body fat from preschool to early adolescence: The Framingham Children’s Study. Int J Obes, 2003;27:827-833.

Rashad I. Structural Estimation of Caloric Intake, Exercise, Smoking, and Obesity.: NBER Work Pap No. 11957, 2006.

Richards TJ, Patterson PM, Tegene A. Obesity and Nutrient Consumption: A Rational Addiction? Morrison School of Agribusiness and Resource Management, 2004.

Robinson T. Does Television Cause Childhood Obesity? JAMA 1998;279:959-960.

Saelen BE, Sallis JF, Frank LD. Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures. Ann Behav Med 2003;25(2):80-91.

Tucker L, Bagwell M. Television Viewing and Obesity in Adult Females. Am J Public Health 1991;81(7):908-911.

Tucker L, Friedman G. Television Viewing and Obesity in Adult Males. Am J Public Health 1989;79(4):516-518.

U.S. Department of Health and Human Services. Physical Activity and Older Americans: Benefits and Strategies. Agency for Healthcare Research and Quality and the Centers for Disease Control 2002.

Wada R, Tekin E. Body Composition and Wages. NBER Work Pap No. 13595, 2007

Wolf A., Colditz G. Current Estimates of the Economic Costs of Obesity in the United States. Obes Res 1998;6:97-106.

World Health Organization. Obesity and Overweight. World Health Organization, 2004.

World Health Organization. Obesity: Preventing and Managing the Global Epidemic. Report of a WHO Consultation on Obesity 1997.

Yach D, Stuckler D, Brownell KD. Epidemiology and economic consequences of the global epidemics of obesity and diabetes. Nat Med 2006;62-66.

Young L, Nestle M. The Contribution of Expanding Portion Sizes to the U.S. obesity epidemic. Am J Public Health 2002;92(2):246-249.

VI. Appendix – Tables

Table 15: Complete list of raw-data

MMSA_Name	MMSA _Code	Spawl _index	**VAR_7 22211**	**VAR _722 110**	**VAR_NP 722211**	**VAR_NP 722110**	restau rants	population02
Akron, OH Metropolitan Statistical Area	10420	105.88	523	418	0	64	1005	699,688
Albuquerque, NM	10740	124.45	529	460	81	68	1138	753,834
Atlanta-Sandy Springs-Marietta, GA Metropolitan Statistical Area	12060	57.66	3,215	3,239	717	706	7877	4,559,952
Baltimore-Towson, MD Metropolitan Statistical Area	12580	115.86	1,881	1,286	422	166	3755	2,595,545
Baton Rouge, LA Metropolitan Statistical Area	12940	90.13	451	341	0	0	792	713,608
Birmingham-Hoover, AL Metropolitan Statistical Area	13820	87.97	754	554	0	126	1434	1,064,511
Boston-Quincy, MA Metropolitan Division	14484	126.93	1,299	1,549	250	156	3254	1,845,554
Cambridge-Newton-Framingham, MA Metropolitan Division	15764	126.93	975	1,013	206	113	2307	1,473,811
Essex County, MA Metropolitan Division	21604	126.93	526	535	119	65	1245	733,688
Bridgeport-Stamford-Norwalk, CT Metropolitan Statistical Area	14860	68.39	503	764	150	93	1510	892,470
Chicago-Naperville-Joliet, IL Metropolitan Division	16974	121.2	5,112	4,526	1,157	821	11616	7,757,890
Cincinnati-Middletown, OH-KY-IN Metropolitan Statistical Area	17140	96.04	1,393	1,259	0	103	2755	2,050,677
Columbia, SC Metropolitan Statistical Area	17900	94.17	432	430	0	0	862	663,535
Dallas-Plano-Irving, TX Metropolitan Division	19124	78.26	2,464	2,155	624	584	5827	3,667,356
Denver-Aurora, CO Metropolitan Statistical Area	19740	125.25	1,552	1,667	263	192	3674	2,280,028
Detroit-Warren-Livonia, MI Metropolitan Statistical Area	19804	79.47	2,846	2,602	429	340	6217	4,488,421
Winston-Salem, NC Metropolitan Statistical Area	49180	46.78	289	316	30	46	681	432,858
Greensboro--Winston-Salem--High Point, NC Combined Statistical Area	24660	46.78	852	965	97	132	2046	657,238
Honolulu, HI Metropolitan Statistical Area	26180	140.21	836	683	157	65	1741	883,357
Houston-Baytown-Sugar Land, TX Metropolitan Statistical Area	26420	93.3	3,049	2,535	935	765	7284	4,988,369
Indianapolis, IN Metropolitan Statistical Area	26900	93.73	1,248	932	107	109	2396	1,578,239
Jacksonville, FL Metropolitan Statistical Area	27260	91.58	786	777	165	202	1930	1,173,245
Kansas City, MO-KS Metropolitan Statistical Area	28140	91.64	1,269	1,165	175	123	2732	1,887,074
Las Vegas-Paradise, NV Metropolitan Statistical Area	29820	104.74	1,054	781	0	0	1835	1,516,604
Little Rock-North Little Rock, AR Metropolitan Statistical	30780	82.27	434	366	59	72	931	623,061
Los Angeles-Long Beach-Glendale, CA Metropolitan Division	31084	101.79	6,516	5,934	2,358	1,701	16509	9,755,021
Memphis, TN-MS-AR Metropolitan Statistical Area	32820	92.15	930	600	161	150	1841	1,224,260
Miami-Miami Beach-Kendall, FL Metropolitan Division	33124	125.68	1,195	1,507	516	709	3927	2,308,355
Milwaukee-Waukesha-West Allis, WI Metropolitan Statistical Area	33340	117.29	857	957	77	86	1977	1,519,705
Minneapolis-St. Paul-Bloomington, MN-WI Metropolitan Statistical Area	33460	95.86	1,789	1,809	158	136	3892	3,055,619
New Haven-Milford, CT Metropolitan Statistical Area	35300	106.97	523	648	202	107	1480	833,105
New Orleans-Metairie-Kenner, LA Metropolitan Statistical Area	35380	125.39	808	968	183	0	1959	1,310,048
Virginia Beach-Norfolk-Newport News, VA-NC Metropolitan Statistical Area	47260	95.63	1,138	1,145	151	133	2567	1,605,420
Omaha-Council Bluffs, NE-IA Metropolitan Statistical Area	36540	128.35	632	471	0	0	1103	782,765
Orlando, FL Metropolitan Statistical Area	36740	96.39	1,127	1,203	261	0	2591	1,757,576
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metropolitan Statistical Area	37964	112.61	3,528	3,523	1,590	446	9087	5,732,438
Phoenix-Mesa-Scottsdale, AZ Metropolitan Statistical Area	38060	110.93	2,295	1,888	313	277	4773	3,497,668

Pittsburgh, PA Metropolitan Statistical Area	38300	105.94	1,473	1,647	322	0	3442	2,410,093
Portland-Vancouver-Beaverton, OR-WA Metropolitan Statistical Area	38900	126.12	1,394	1,603	284	156	3437	2,014,037
Durham, NC Metropolitan Statistical Area	20500	54.17	319	308	42	0	669	442,534
Ogden-Clearfield, UT Metropolitan Statistical Area	36260	110.92	293	0	28	0	321	461,981
Salt Lake City, UT Metropolitan Statistical Area	41620	110.92	776	565	106	82	1529	1,003,505
Seattle-Bellevue-Everett, WA Metropolitan Division	42644	100.91	1,777	2,099	383	232	4491	2,395,556
Springfield, MA Metropolitan Statistical Area	44140	122.49	422	506	107	68	1103	682,749
St. Louis, MO-IL Metropolitan Statistical Area	41180	94.51	1,864	1,624	0	0	3488	2,734,424
Tacoma, WA Metropolitan Division	45104	105.88	407	408	105	47	967	729,518
Tampa-St. Petersburg-Clearwater, FL Metropolitan Statistical Area	45300	86.26	1,265	1,671	516	477	3929	2,485,951
Toledo, OH Metropolitan Statistical Area	45780	107.19	479	509	0	47	1035	658,610
Tucson, AZ Metropolitan Statistical Area	46060	109.13	575	511	86	64	1236	877,666
Tulsa, OK Metropolitan Statistical Area	46140	99.06	696	547	0	0	1243	877,410
Washington-Arlington-Alexandria, DC-VA-MD-WV Metropolitan Statistical Area	47894	90.83	3,573	3,298	622	457	7950	5,015,705
Wichita, KS Metropolitan Statistical Area	48620	110.09	482	367	0	0	849	579,024
Worcester, MA Metropolitan Statistical Area	49340	90.48	463	495	0	70	1028	767,835

Table 16: NHWM

Weight Model, Adjustment Group: Non-Hispanic White, Male

The REG Procedure
Model: PredictedWeight
Dependent Variable: BMXWT Weight (kg)

Number of Observations Read	1233
Number of Observations Used	1233

Note: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	9654223	4827111	301363	<.0001
Error	1231	19718	16.01759		
Uncorrected Total	1233	9673940			

Root MSE	4.00220	**R-Square**	0.9980
Dependent Mean	86.69181	**Adj R-Sq**	0.9980
Coeff Var	4.61658		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
WHD020	Current Self-reported weight (kg)	1	0.99855	0.00577	173.05	<.0001
WHD020_sqrd	Current Self-reported weight (kg^2) Squared	1	-0.00000850	0.00005947	-0.14	0.8864

Table 17: NHWF

Weight Model, Adjustment Group: Non-Hispanic White, Female

The REG Procedure
Model: PredictedWeight
Dependent Variable: BMXWT Weight (kg)

Number of Observations Read	1329
Number of Observations Used	1329

Note: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	7441553	3720777	150254	<.0001
Error	1327	32861	24.76330		
Uncorrected Total	1329	7474414			

Root MSE	4.97627	**R-Square**	0.9956
Dependent Mean	72.82370	**Adj R-Sq**	0.9956
Coeff Var	6.83332		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
WHD020	Current Self-reported weight (kg)	1	1.04676	0.00773	135.38	<.0001
WHD020_sqrd	Current Self-reported weight (kg^2) Squared	1	-0.00019439	0.00009480	-2.05	0.0405

Table 18: NHBM

Weight Model, Adjustment Group: Non-Hispanic Black, Male

The REG Procedure
Model: PredictedWeight
Dependent Variable: BMXWT Weight (kg)

Number of Observations Read	518
Number of Observations Used	518

Note: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3941969	1970984	65639.9	<.0001
Error	516	15494	30.02721		
Uncorrected Total	518	3957463			

Root MSE	5.47971	R-Square	0.9961
Dependent Mean	84.92954	Adj R-Sq	0.9961
Coeff Var	6.45207		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
WHD020	Current Self-reported weight (kg)	1	0.93083	0.01202	77.42	<.0001
WHD020_sqrd	Current Self-reported weight (kg^2) Squared	1	0.00064095	0.00012357	5.19	<.0001

Table 19: NHBF

Weight Model, Adjustment Group: Non-Hispanic Black, Female

The REG Procedure
Model: PredictedWeight
Dependent Variable: BMXWT Weight (kg)

Number of Observations Read	537
Number of Observations Used	537

Note: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3648419	1824210	58437.9	<.0001
Error	535	16701	31.21623		
Uncorrected Total	537	3665120			

Root MSE	5.58715	R-Square	0.9954
Dependent Mean	79.78529	Adj R-Sq	0.9954
Coeff Var	7.00273		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
WHD020	Current Self-reported weight (kg)	1	1.04576	0.01069	97.84	<.0001
WHD020_sqrd	Current Self-reported weight (kg^2) Squared	1	-0.00022731	0.00011488	-1.98	0.0484

Table 20: HM

Weight Model, Adjustment Group: Hispanic, Male

The REG Procedure
Model: PredictedWeight
Dependent Variable: BMXWT Weight (kg)

Number of Observations Read	628
Number of Observations Used	628

Note: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	2	4170733	2085366	79810.3	<.0001
Error	626	16357	26.12904		
Uncorrected Total	628	4187089			

Root MSE	5.11166	**R-Square**	0.9961
Dependent Mean	79.68360	**Adj R-Sq**	0.9961
Coeff Var	6.41494		

Parameter Estimates						
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
WHD020	Current Self-reported weight (kg)	1	0.94894	0.01084	87.50	<.0001
WHD020_sqrd	Current Self-reported weight (kg^2) Squared	1	0.00058188	0.00012095	4.81	<.0001

Table 21: HF

Weight Model, Adjustment Group: Hispanic, Female

The REG Procedure
Model: PredictedWeight
Dependent Variable: BMXWT Weight (kg)

Number of Observations Read	662
Number of Observations Used	662

Note: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	2	3462129	1731065	64954.9	<.0001
Error	660	17589	26.65025		
Uncorrected Total	662	3479718			

Root MSE	5.16239	**R-Square**	0.9949
Dependent Mean	70.61767	**Adj R-Sq**	0.9949
Coeff Var	7.31033		

Parameter Estimates						
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
WHD020	Current Self-reported weight (kg)	1	1.04709	0.01200	87.28	<.0001
WHD020_sqrd	Current Self-reported weight (kg^2) Squared	1	-0.00026799	0.00015296	-1.75	0.0802

Table 22: MRM

Weight Model, Adjustment Group: Other Race - Including Multi-Racial, Male

The REG Procedure
Model: PredictedWeight
Dependent Variable: BMXWT Weight (kg)

Number of Observations Read	78
Number of Observations Used	78

Note: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	479189	239594	21815.0	<.0001
Error	76	834.70730	10.98299		
Uncorrected Total	78	480023			

Root MSE	3.31406	R-Square	0.9983
Dependent Mean	76.24359	Adj R-Sq	0.9982
Coeff Var	4.34667		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
WHD020	Current Self-reported weight (kg)	1	0.95612	0.01897	50.40	<.0001
WHD020_sqrd	Current Self-reported weight (kg^2) Squared	1	0.00059623	0.00021680	2.75	0.0074

Table 23: MRF

Weight Model, Adjustment Group: Other Race - Including Multi-Racial, Female

The REG Procedure
Model: PredictedWeight
Dependent Variable: BMXWT Weight (kg)

Number of Observations Read	99
Number of Observations Used	99

Note: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	433168	216584	8421.00	<.0001
Error	97	2494.79369	25.71952		
Uncorrected Total	99	435663			

Root MSE	5.07144	R-Square	0.9943
Dependent Mean	64.59192	Adj R-Sq	0.9942
Coeff Var	7.85151		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
WHD020	Current Self-reported weight (kg)	1	1.01385	0.03517	28.83	<.0001
WHD020_sqrd	Current Self-reported weight (kg^2) Squared	1	0.00042230	0.00050129	0.84	0.4016

Table 24: NHWM

Height Model, Adjustment Group: Non-Hispanic White, Male

The REG Procedure
Model: PredictedHeight
Dependent Variable: BMXHT Standing Height (cm)

Number of Observations Read	1233
Number of Observations Used	1233

Note: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	38456276	19228138	1816679	<.0001
Error	1231	13029	10.58422		
Uncorrected Total	1233	38469305			

Root MSE	3.25334	R-Square	0.9997
Dependent Mean	176.47802	Adj R-Sq	0.9997
Coeff Var	1.84348		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
WHD010	Current Self-reported height (cm)	1	1.07720	0.01258	85.66	<.0001
WHD010_sqrd	Current Self-reported height (cm^2) Squared	1	-0.00047969	0.00007035	-6.82	<.0001

Table 25: NHWF

Height Model, Adjustment Group: Non-Hispanic White, Female

The REG Procedure
Model: PredictedHeight
Dependent Variable: BMXHT Standing Height (cm)

Number of Observations Read	1329
Number of Observations Used	1329

Note: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	35195470	17597735	2754688	<.0001
Error	1327	8477.25526	6.38829		
Uncorrected Total	1329	35203947			

Root MSE	2.52751	R-Square	0.9998
Dependent Mean	162.61309	Adj R-Sq	0.9998
Coeff Var	1.55431		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
WHD010	Current Self-reported height (cm)	1	1.07449	0.01013	106.03	<.0001
WHD010_sqrd	Current Self-reported height (cm^2) Squared	1	-0.00048268	0.00006177	-7.81	<.0001

Table 26: NHBM

Height Model, Adjustment Group: Non-Hispanic Black, Male

The REG Procedure
Model: PredictedHeight
Dependent Variable: BMXHT Standing Height (cm)

Number of Observations Read	518
Number of Observations Used	518

Note: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	16261034	8130517	745448	<.0001
Error	516	5627.95731	10.90689		
Uncorrected Total	518	16266662			

Root MSE	3.30256	R-Square	0.9997
Dependent Mean	177.07239	Adj R-Sq	0.9997
Coeff Var	1.86509		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
WHD010	Current Self-reported height (cm)	1	1.18710	0.01901	62.46	<.0001
WHD010_sqrd	Current Self-reported height (cm^2) Squared	1	-0.00109	0.00010608	-10.25	<.0001

Table 27: NHBF

Height Model, Adjustment Group: Non-Hispanic Black, Female

The REG Procedure
Model: PredictedHeight
Dependent Variable: BMXHT Standing Height (cm)

Number of Observations Read	537
Number of Observations Used	537

Note: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	14269236	7134618	730785	<.0001
Error	535	5223.17990	9.76295		
Uncorrected Total	537	14274459			

Root MSE	3.12457	R-Square	0.9996
Dependent Mean	162.91210	Adj R-Sq	0.9996
Coeff Var	1.91795		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
WHD010	Current Self-reported height (cm)	1	1.19083	0.01914	62.21	<.0001
WHD010_sqrd	Current Self-reported height (cm^2) Squared	1	-0.00119	0.00011635	-10.27	<.0001

Table 28: HM

Height Model, Adjustment Group: Hispanic, Male

The REG Procedure
Model: PredictedHeight
Dependent Variable: BMXHT Standing Height (cm)

Number of Observations Read	628
Number of Observations Used	628

Note: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	18252697	9126348	575999	<.0001
Error	626	9918.57739	15.84437		
Uncorrected Total	628	18262616			

Root MSE	3.98050	R-Square	0.9995
Dependent Mean	170.38073	Adj R-Sq	0.9995
Coeff Var	2.33624		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
WHD010	Current Self-reported height (cm)	1	1.26969	0.01931	65.75	<.0001
WHD010_sqrd	Current Self-reported height (cm^2) Squared	1	-0.00160	0.00011220	-14.23	<.0001

Table 29: HF

Height Model, Adjustment Group: Hispanic, Female

The REG Procedure
Model: PredictedHeight
Dependent Variable: BMXHT Standing Height (cm)

Number of Observations Read	662
Number of Observations Used	662

Note: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	16432739	8216370	424877	<.0001
Error	660	12763	19.33825		
Uncorrected Total	662	16445502			

Root MSE	4.39753	R-Square	0.9992
Dependent Mean	157.48082	Adj R-Sq	0.9992
Coeff Var	2.79242		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
WHD010	Current Self-reported height (cm)	1	1.33293	0.02323	57.39	<.0001
WHD010_sqrd	Current Self-reported height (cm^2) Squared	1	-0.00213	0.00014579	-14.64	<.0001

Table 30: MRM

Height Model, Adjustment Group: Other Race - Including Multi-Racial, Male

The REG Procedure
Model: PredictedHeight
Dependent Variable: BMXHT Standing Height (cm)

Number of Observations Read	78
Number of Observations Used	78

Note: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2298232	1149116	225231	<.0001
Error	76	387.74761	5.10194		
Uncorrected Total	78	2298619			

Root MSE	2.25875	R-Square	0.9998
Dependent Mean	171.53462	Adj R-Sq	0.9998
Coeff Var	1.31679		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
WHD010	Current Self-reported height (cm)	1	1.11543	0.03477	32.08	<.0001
WHD010_sqrd	Current Self-reported height (cm^2) Squared	1	-0.00068569	0.00020122	-3.41	0.0011

Table 31: MRF

Height Model, Adjustment Group: Other Race - Including Multi-Racial, Female

The REG Procedure
Model: PredictedHeight
Dependent Variable: BMXHT Standing Height (cm)

Number of Observations Read	99
Number of Observations Used	99

Note: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2476483	1238241	193068	<.0001
Error	97	622.10861	6.41349		
Uncorrected Total	99	2477105			

Root MSE	2.53249	R-Square	0.9997
Dependent Mean	158.03737	Adj R-Sq	0.9997
Coeff Var	1.60246		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
WHD010	Current Self-reported height (cm)	1	1.03838	0.03821	27.17	<.0001
WHD010_sqrd	Current Self-reported height (cm^2) Squared	1	-0.00028159	0.00023918	-1.18	0.2420

Table 32: SAS Adjusted R-square

The SAS System

The REG Procedure
Model: PredictedCommunityBMI
Dependent Variable: Mean

Adjusted R-Square Selection Method

Number of Observations Read	43
Number of Observations Used	43

Number in Model	Adjusted R-Square	R-Square	Variables in Model
4	0.2160	0.2907	VAR_722211_cap VAR_722110_cap interaction2 interaction3
5	0.1952	0.2910	Spawl_index VAR_722211_cap VAR_722110_cap interaction2 interaction3
1	0.1191	0.1401	interaction2
3	0.1159	0.1791	VAR_722211_cap VAR_722110_cap interaction3
2	0.1125	0.1547	Spawl_index VAR_722110_cap
3	0.1019	0.1661	Spawl_index VAR_722211_cap VAR_722110_cap
2	0.0985	0.1415	VAR_722110_cap interaction2
2	0.0980	0.1410	VAR_722211_cap interaction2
2	0.0972	0.1402	Spawl_index interaction2
2	0.0971	0.1401	interaction2 interaction3
4	0.0952	0.1814	Spawl_index VAR_722211_cap VAR_722110_cap interaction3
3	0.0923	0.1571	Spawl_index VAR_722110_cap interaction3
3	0.0902	0.1551	Spawl_index VAR_722110_cap interaction2
4	0.0873	0.1742	Spawl_index VAR_722211_cap VAR_722110_cap interaction2
3	0.0801	0.1458	VAR_722211_cap VAR_722110_cap interaction2
3	0.0759	0.1419	VAR_722110_cap interaction2 interaction3
3	0.0750	0.1411	VAR_722211_cap interaction2 interaction3
3	0.0749	0.1410	Spawl_index VAR_722211_cap interaction2
3	0.0742	0.1404	Spawl_index interaction2 interaction3
4	0.0695	0.1581	Spawl_index VAR_722110_cap interaction2 interaction3
2	0.0663	0.1107	VAR_722110_cap interaction3
4	0.0516	0.1419	Spawl_index VAR_722211_cap interaction2 interaction3

The SAS System

The REG Procedure
Model: PredictedCommunityBMI
Dependent Variable: Mean

Adjusted R-Square Selection Method

Number in Model	Adjusted R-Square	R-Square	Variables in Model
1	0.0358	0.0587	VAR_722110_cap
1	0.0341	0.0571	Spawl_index
2	0.0333	0.0793	VAR_722211_cap VAR_722110_cap
1	0.0297	0.0528	interaction3
2	0.0159	0.0628	Spawl_index VAR_722211_cap
2	0.0144	0.0613	Spawl_index interaction3
2	0.0102	0.0574	VAR_722211_cap interaction3
3	-.0092	0.0629	Spawl_index VAR_722211_cap interaction3
1	-.0241	0.0003	VAR_722211_cap

Table 33: Model for Total Restaurants

The REG Procedure
Model: PredictedCommunityBMI
Dependent Variable: Mean

Number of Observations Read	43
Number of Observations Used	43

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1.44711	0.48237	1.78	0.1662
Error	39	10.54595	0.27041		
Corrected Total	42	11.99306			

Root MSE	0.52001	R-Square	0.1207
Dependent Mean	27.34920	Adj R-Sq	0.0530
Coeff Var	1.90137		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	28.43922	1.35678	20.96	<.0001
Spawl_index	Spawl_index	1	-0.00295	0.01566	-0.19	0.8515
Total_Per_Cap_Restaurants	Total PC # Restaurants	1	-257.00373	798.99378	-0.32	0.7494
interaction_total	Total PC # Restaurants x Sprawl Interaction	1	-3.25564	9.97936	-0.33	0.7460

VII. **Verzeichnis der akademischen Lehrer**

Meine akademischen Lehrer waren Damen/Herren in Marburg; Adler, Amon, Baum, Basler, Becker, Berndt, Blankenburg, Brandis, Drenckhan, Eschenbach, Friedrich, Fruhstorfer, Gressner, Griss , Huffmann, Ihm, Kern, Kleinsasser, Lennartz, Mannheim, Müller, Oepen, Portig, Rodeck, Schaumlöffel, Schmitt, Schulz, Schüffel, Siegrist, Thomas, Vohland, Wesemann, Zelder.

VIII. **Danksagung**

I would like to extend my thanks to Prof. Dr.med. Annette Becker M.P.H. who invested her time and gave valuable advice towards the development of this paper. I would also like to extend my deepest appreciation to Anne G. Krausse M.B.A., B.A., for all the advice, encouragement and patience. I am grateful for all her support.